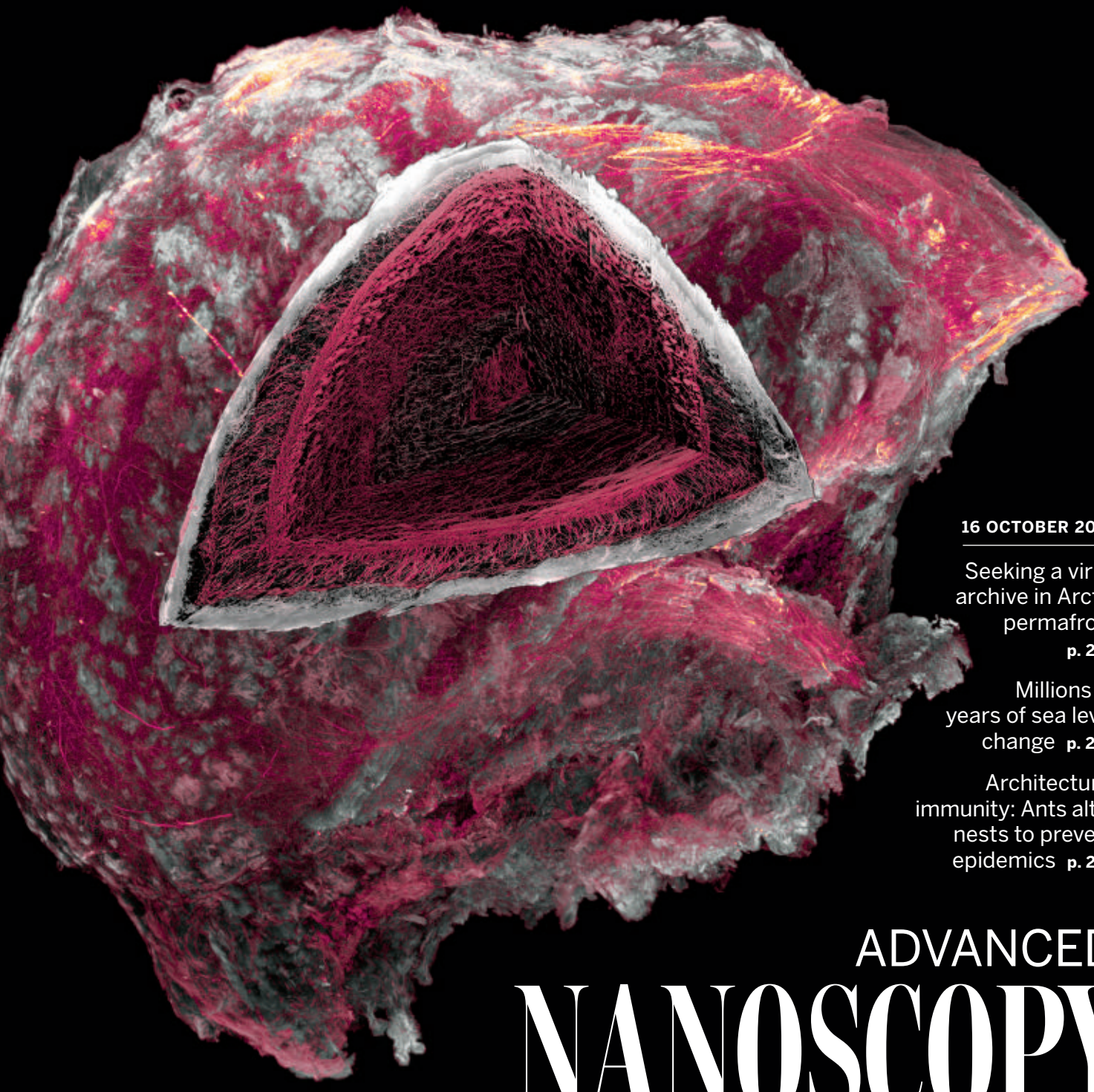
AAAS

Science



16 OCTOBER 2025

Seeking a virus archive in Arctic permafrost **p. 230**

Millions of years of sea level change **p. 260**

Architectural immunity: Ants alter nests to prevent epidemics **p. 266**

ADVANCED NANOSCOPY

3D rendering of photochemically sectioned brain tissue **p. 257**



The Chen Institute and *Science* launched the "Chen Institute and *Science* Prize for AI Accelerated Research" in August 2024. This initiative aims to drive advancements in artificial intelligence that can expand scientific research. Young scientists from around the world are invited to submit AI-driven projects that demonstrate significant potential to improve research and lives.

The competition will award a Grand Prize of $30,000, with the winner's essay published in *Science* and an accompanying five-year AAAS membership. Additionally, up to two runners-up will receive $10,000 each, with their essays published on *Science* Online and the same membership benefits, promoting sustained engagement with scientific progress.

Apply at

cheninstitute.org/prize

CONTENTS

16 OCTOBER 2025 | VOLUME 390 | ISSUE 6770



230

Svalbard's permafrost could preserve RNA viruses deposited by generations of seabirds and other wildlife.

PHOTO: LUKASZ LARSSON WARZECHA

EDITORIAL

217 A view of Horizon Europe's future —L. Borrell-Damián

NEWS

222 Papers based on open health data face bans
PLOS, Frontiers, and others announce policies intended to stem the tide of suspect papers
—C. O'Grady

223 DNA from rum-soaked fishes chronicles century of change
Museum specimens reveal loss of genetic diversity in marine fishes of the Philippines
—E. Stokstad

225 Did lead poisoning help drive human evolution?
"Bold" hypothesis suggests tolerance for lead allowed *Homo sapiens* to outlive Neanderthals
—M. Price

226 Architects of molecular cages win Chemistry Nobel
Spongelike materials called metal-organic frameworks can separate and store gases
—D. Clery and C. Offord

228 Chaos and confusion as U.S. shutdown drags on
Mass layoffs, and subsequent reversals, have added to research spending woes —*Science* News Staff

FEATURES

230 Cold storage
On an Arctic archipelago, frozen soil may preserve a hidden history of viruses —K. Kupferschmidt
PODCAST

COMMENTARY

PERSPECTIVES

236 Kiss, shrink, run
A unified mechanism directs synaptic vesicle release —K. Lichter
RESEARCH SUMMARY p. 258

237 Making hydrogen production durable
An interphase shields polymer electrolytes from electrochemical oxidation in an alkaline environment
—Y. Seung Kim
RESEARCH ARTICLE p. 294

239 Neural compass in the sky
Head-direction neurons maintain stable directional signals during large-scale navigation in the wild —Y.-Q. Zhou and J. J. Knierim
RESEARCH SUMMARY p. 259

241 Quantum trions in equilibrium
Quantum fluid of trions is demonstrated in two-dimensional material systems
—T. Ouyang and S.-F. Shi
RESEARCH ARTICLES pp. 299 & 304

POLICY FORUM

242 Mobilizing capital and technology for a clean aviation industry
Research managers and investors need better approaches to weigh risks and potential benefits of scalable, transformational technologies
—D. G. Victor *et al.*

BOOKS ET AL.

246 Knowledge for two
A psychologist explores common knowledge and coordination
—J. Goodman and C. Firestone

CONTENTS

247 A radically organic worldview turns 100
A philosopher's seminal book traced how science shaped the mentality of the modern world
—À. Gómez-Marín

LETTERS

248 Stemming green tides in the Yellow Sea —L. Yang *et al.*

248 Australian algal blooms require national action
—D. McAfee and B. Martin

249 Chile's framework law erodes institutions
—I. Ojeda-Pereira

RESEARCH

HIGHLIGHTS

251 From *Science* and other journals

RESEARCH SUMMARIES

254 Proteomics
Mapping early human blood cell differentiation using single-cell proteomics and transcriptomics
—B. Furtwängler *et al.*

255 Lineage tracing
High-resolution spatial mapping of cell state and lineage dynamics in vivo with PEtracer
—L. W. Koblan *et al.*

256 Pigmentation
A mosaic of modular variation at a single gene underpins convergent plumage coloration
—D. Lutgen *et al.*

257 Imaging
Mesoscale volumetric fluorescence imaging at nanoscale resolution by photochemical sectioning
—W. Wang *et al.*

258 Cell biology
"Kiss-shrink-run" unifies mechanisms for synaptic vesicle exocytosis and hyperfast recycling
—C.-L. Tao *et al.*
PERSPECTIVE p. 236

259 Neuroscience
Head-direction cells as a neural compass in bats navigating outdoors on a remote oceanic island
—S. Palgi *et al.*
PERSPECTIVE p. 239



260 Global sea level
Global mean sea level over the past 4.5 million years
—P. U. Clark *et al.*

RESEARCH ARTICLES

261 Organic chemistry
The total synthesis of (−)-spiroaspertrione A: A divinylcyclopropane rearrangement approach
—W. Huang *et al.*

266 Infectious disease
Architectural immunity: Ants alter their nest networks to prevent epidemics —L. Leckie *et al.*
PODCAST

272 Organic chemistry
Visible light–driven stereodivergent allylation of cyclic hemiacetals with butene for polypropionate synthesis —H. Nakao *et al.*

279 Animal symbiosis
Defensive fungal symbiosis on insect hindlegs —T. Nishino *et al.*

284 Plant ecology
Drought intensity and duration interact to magnify losses in primary productivity
—T. Ohlert *et al.*

290 Organic chemistry
Synthesis of triple stranded porphyrin nanobelts
—A. Rodríguez-Rubio *et al.*

294 Electrochemistry
Durable, pure water–fed, anion-exchange membrane electrolyzers through interphase engineering —S. Hou *et al.*
PERSPECTIVE p. 237

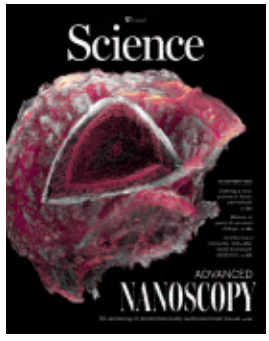
2D materials

299 Electrically controlled interlayer trion fluid in electron-hole bilayers
—R. Qi *et al.*

304 An equilibrium trion liquid in atomic double layers
—P. X. Nguyen *et al.*

PERSPECTIVE p. 241

ON THE COVER



This three-dimensional rendering of a mouse olfactory bulb exposes axons and myelin sheaths, the neuronal wiring that transmits odor information. The dataset was acquired using volumetric imaging via photochemical sectioning (VIPS), which alternates light-driven sectioning and on-block imaging to build a seamless volume. VIPS enables organ-wide nanoscale reconstruction for quantitative mapping of connectivity and molecular landscape in intact tissue. See page 257. Credit: Gao laboratory, University of Illinois Chicago; Upadhyayula laboratory, University of California, Berkeley

ON THE PODCAST

Hunting ancient viruses in the Arctic, and how ants build their nests to fight disease

WORKING LIFE

310 An immigrant's dilemma
—M. Gelambi

308 Science Careers

Science **serves as a forum for discussion of important issues related to the advancement of science by publishing material on which a consensus has been reached as well as including the presentation of minority or conflicting points of view. Accordingly, all articles published in *Science*—including editorials, news, commentary, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.** *Science* (ISSN 0036-8075) is published weekly on Thursday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices.

Copyright © 2025 by the American Association for the Advancement of Science. The title *Science* is a registered trademark of the AAAS.

Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $2865; Foreign postage extra: Air assist delivery: $135. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A. Change of address:** Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling available from backissues.science.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

PHOTO: YUVAL BARKAI

A view of Horizon Europe's future

Lidia Borrell-Damián

The European Union (EU) has the world's largest funding program exclusively for research and innovation (R&I)—Horizon Europe—which is now preparing its 10th edition for 2028–2034. The good news is that the EU aims to raise Horizon Europe's budget by 83% and continue the trend of supporting all types of institutions and partnerships, including universities, research institutes, small- and medium-sized enterprises, nonprofit organizations, and a widening circle of international collaborations.

Horizon Europe is the nickname given to both the 9th and 10th EU Framework Programmes for Research and Innovation, with goals that require a careful balance among long- and short-term gains, security and openness, and economic growth and social well-being. R&I have the potential, however, to bridge these seemingly competing objectives. A key factor to success will be the budget: The proposed allocation of €175 billion is a welcome and necessary increase over the €95 billion budget for the 9th Framework Programme for R&I. Although less than the €200 billion that was recommended by the research sector to achieve the program's objectives, it nevertheless forms a solid foundation for the future. It is now crucial to ensure that this funding is maintained during the upcoming budget negotiations by the EU institutions at play: namely, the European Commission (creating the initial proposals), the Council of the EU (member state ministers), and the European Parliament (elected representatives on the EU level). Additionally, structures should be put in place to protect the program's legal distinction from other EU funding programs and to prevent its funding from being reallocated to non-R&I priorities.

...research is a core component of competitiveness, not a tool in its service.

Horizon Europe will be structured across four "pillars" (Excellent Science; Competitiveness and Society; Innovation; and European Research Area) for continuity with previous programs. The first pillar features a bottom-up approach to research that champions fundamental science and boasts prestigious instruments such as the European Research Council and the Marie Skłodowska-Curie Actions, which support researchers' careers and training. These will continue to fund projects that embrace scientific curiosity and freedom of inquiry exclusively on the basis of scientific excellence. Safeguarding their governance and independence from policy-driven objectives is essential to maintaining Europe as a leader in bringing about new ideas and innovative solutions.

The proposed approach to competitiveness, however, gives some cause for concern. A focus on short-term policy goals, such as rapid economic growth, risks reducing the scope for bottom-up approaches to research. The new Framework Programme is meant to coordinate with the European Commission's European Competitiveness Fund, a merger of multiple funding initiatives primarily focused on scaling up the EU's industrial and economic development. This might divert Horizon funds from fundamental, long-term research toward priorities dominated by short-term competitiveness criteria. It would be more advantageous for Horizon Europe to uphold a long-term definition of competitiveness in which research is a core component of competitiveness, not a tool in its service.

Horizon Europe's continued commitment to reduce persisting disparities in R&I capacity across Europe through its "Widening" program promises a more integrated and cohesive European R&I environment. Here, more focus should be placed on "brain circulation" to increase the research capacity of regions and countries that lag behind to achieve a level playing field across the continent. However, clearer action is needed to reduce the gap in R&I performance between European countries.

In a time of international political conflicts, science is a bridge for dialogue and cooperation. Although strategic areas such as defense and advanced technology may require special security measures, these must not create arbitrary barriers to scientific openness and collaboration. Association with like-minded countries outside the EU and opportunities for researchers in the EU's "Choose Europe" initiative are good examples of how the EU is supporting international collaboration while protecting researchers at risk from political pressure or threats to academic freedom. In 2025, Switzerland and South Korea officially became fully associated with Horizon Europe, and Egypt is expected to do so as well later this year. Japan is likely to join in 2026, and Australia has recently began exploring a potential association agreement.

There are other important issues in Horizon Europe that require more detailed consideration, such as administrative simplification, environmentally sustainable research practices, and inclusivity. However, the Horizon Europe proposal builds on the legacy of previous programs and on Europe's strong research traditions, while expanding its global reach and ambition. □

Lidia Borrell-Damián is the secretary general of Science Europe, Brussels, Belgium. lidia.borrell-damian@scienceeurope.org

10.1126/science.aed0012



Unmet needs, untapped opportunities: Attention on women's health

Science and business strategy are joining forces to dismantle the historical neglect of female biology, leading to the development of more precise, equitable care and a new generation of solutions to improve women's health.

As the world slowed during the pandemic, biochemist Ian Cottingham retreated to the peace of his home with a new sense of purpose. After a decade as research director at PPL Therapeutics, the company that cloned Dolly the sheep, and 16 years leading drug development at Ferring Pharmaceuticals, active in reproductive medicine and maternal health, he was ready for a different challenge as a startup founder. He chose to focus on preeclampsia, a deadly pregnancy complication medicine has failed to address. "No new treatments have been approved for preeclampsia," he says.

For centuries, medicine operated under a flawed premise, "using the male body as the standard for most of our medical knowledge," explains Eva Hoffmann, a professor of molecular genetics and the prorector for research and innovation at the University of Copenhagen. This systemic bias, she notes, "has really limited our ability to understand and therefore give proper diagnosis and treatment for women."

Such bias has left major gaps in knowledge—not only in reproductive health and conditions unique to women, but also in diseases that affect the sexes differently, such as cardiovascular diseases and Alzheimer's. These gaps mean diagnostics and treatments may be less effective for women, if they are available at all—there may even be a higher risk of adverse reactions due to biological differences.

Today, collaborations at the intersection of science and business are helping to break down the long-standing barriers that have kept female biology overlooked and underserved. This transformation is being championed by institutions like the BioInnovation Institute (BII), which strategically fosters promising start-ups and company creation projects, including in areas that have traditionally received little attention. "A central part of

IMAGE: © AYA_LINES_DESIGN/SHUTTERSTOCK.COM

BII's mission is to tackle pressing health challenges," says Jens Nielsen, CEO of BII. "Women's health is a clear example, since the field still faces major gaps in medical innovation." By building an innovation ecosystem that goes far beyond providing funding, BII is helping to power the development of a new generation of solutions to improve health outcomes worldwide.

Anatomy of neglect



Preeclampsia is distinctive in the history of medicine for being understudied despite its severity. Preeclampsia is a hypertensive disorder unique to pregnancy that can lead to seizures, stroke, organ failure, and death for the mother, while starving the developing fetus of oxygen and forcing a dangerously premature birth. The standard of care—blood pressure medications that prolong pregnancy, on average, just 10 extra days—ultimately ends with premature delivery to save the mother, shifting risk to the newborn.

Every year, preeclampsia is responsible for the deaths of 70,000 women and 500,000 infants globally. Beyond the immediate fatalities, an estimated 300 million people are believed to suffer from long-term health consequences stemming from the condition, such as increased risk of cardiovascular disease, chronic kidney disease, and type 2 diabetes for mothers, and neurodevelopmental disorders and metabolic issues for children. In the U.S. alone, out of approximately 3.8 million annual pregnancies, there are about 140,000 cases of preeclampsia, with 40,000 of those being severe. The economic burden is also staggering; while there is limited data on global costs, according to a report from 2017, the short-term costs of preeclampsia to the U.S. healthcare system are $2.18 billion.

Why has such a deadly, costly, and common condition been so thoroughly ignored by medical research and development? Cottingham points to a confluence of factors rooted in systemic bias. First, the very nature of pregnancy may make the problem seem temporary, a transient crisis that ends with birth, thus failing to attract the long-term research and development investment seen in chronic diseases. Second, and most critically, the immense challenge of ensuring fetal safety has created a chilling effect on innovation. The placenta, a complex and dynamic organ, acts as both a gateway and a barrier, and the risk of a new drug crossing it and harming the fetus has been a deterrent to progress for generations, Cottingham explains.

> **"If we want momentum, we need to help create more visible wins."**
>
> **—Jens Nielsen**

It is precisely this challenge—the placental barrier—that Cottingham, now CEO and co-founder of Beech Biotech, decided to tackle head-on. The core of the company's research is a new delivery mechanism born of elegant bioengineering. "We came up with the idea of making antibodies that didn't cross the placenta," he says. "And that gives you all of the flexibility of antibodies, but without the major concerns about fetal safety."

Protecting babies, treating mothers

During a healthy pregnancy, the placenta releases a protein receptor called fms-like tyrosine kinase-1 (Flt-1), which acts like a docking station for two crucial growth factors circulating in the blood: vascular endothelial growth factor (VEGF) and placental growth factor (PlGF). When VEGF and PlGF bind to the Flt-1 receptor on a cell, they signal it to maintain the health and integrity of the blood vessels, which is especially important for the development of the placenta.

In preeclampsia, the placenta produces an excess of an altered, shorter version of the Flt-1 receptor. This version doesn't stay on the cell surface. Instead, it is released and circulates freely in the mother's bloodstream. It acts like a sponge, intercepting and binding to VEGF and PlGF before they can reach their intended receptors on the blood vessel walls, effectively starving the mother's blood vessels of the signals they

PHOTO: PROVIDED BY ESBEN ZØLLNER OLESEN/BIOINNOVATION INSTITUTE.

need to stay healthy. Beech Biotech has designed its therapeutic antibody to bind to and clear excess Flt-1, restoring balance and allowing the pregnancy to continue safely. They have cleverly modified the antibody so that it becomes effectively confined to the mother's circulation, says Cottingham, with the aim to "take the potential risks of fetal harm out of the equation." Beech Biotech's recent studies in non-human primates support the potential of the approach.

Focus on biology and the path forward

Beech Biotech's targeted approach to preeclampsia is a powerful example of a much broader, more profound shift occurring in medicine. For too long, women's health has been treated as a narrow, siloed discipline, focused largely on female oncology or fertility. The rest of the female body was often treated as a variation of the male default, Hoffmann says. "Researchers in women's health have known the field is deeply underfunded, both for basic research and understanding of the biological reasons for female-specific conditions," she says. "For those working on topics like menstruation, endometriosis, pregnancy, or menopause, it has been hard at times to be taken seriously," as these conditions were often brushed aside with the logic that "Well, that's just what it means to be a woman."

Nielsen points to long-standing systemic barriers: decades of limited investment—just 1% of biopharma R&D in 2020 went to female-specific conditions outside cancer—led to fewer studies and data gaps. Stigma, lack of awareness, and limited commercial success have further slowed progress. "If we want momentum, we need to help create more visible wins," Nielsen says.

While he stresses that innovation in women's health is still at an early stage compared to more established areas, Nielsen notes that growing attention—driven by advocates, researchers, and entrepreneurs—is beginning to change the trajectory. There is growing recognition that sex is a fundamental biological variable influencing how and when diseases appear, progress, and respond to treatment across every system of the body. For instance, Alzheimer's disease afflicts women at nearly twice the rate of men, a discrepancy that cannot be explained by



Ian Cottingham's company Beech Biotech, with support from BII, is developing treatments for preeclampsia, a potentially deadly complication of pregnancy.

PHOTO: PROVIDED BY ESBEN ZØLLNER OLESEN/BIOINNOVATION INSTITUTE.

differences in life expectancy. In cardiovascular disease, studied primarily in men for decades, we now know that women can present with different heart attack symptoms, such as nausea and jaw pain, rather than the classic chest-clutching, potentially leading to fatal misdiagnoses if these symptoms are missed.

The underlying pathology and risk factors can also differ significantly. Hoffmann says women are more prone to conditions like small vessel disease and non-obstructive coronary artery disease. "We can't forget that cardiovascular disease is a leading cause of death in women," Hoffmann says. And yet, she argues, the gap in understanding is "very symptomatic of the misconception that you can assume women are just small men."

Researchers are now investigating the complex interplay of genetics, hormones, and immune responses to understand women's specific disease vulnerabilities. This recalibration in science is revealing fundamental sex-based differences in everything from disease patterns—like the higher prevalence of autoimmune conditions in women—to basic pharmacology, including the different ways drugs are metabolized.

According to Hoffmann, driving this transformation is a combination of societal evolution and scientific proof-of-concept. The success of focused campaigns, like those for breast cancer, which saw survival rates jump from 50% to 90% over the last two decades, demonstrated that "efforts in women's health really matter and can make a huge difference," she says. Most importantly, societal changes have driven change within the scientific community. In addition to having a greater presence in research labs, women in science are slowly ascending to senior leadership and decision-making positions, which means that those directly affected by these health gaps now have agency in setting research strategies and deciding where funds are allocated.

The catalyst to progress

A brilliant scientific idea alone, even one as compelling as Beech Biotech's, is not enough to survive the perilous journey from the lab to the market. For women's health innovation in particular, ecosystem support is crucial. "It's a young field that needs not just financing, but also networks, expertise, and strategic guidance to mature," Nielsen adds. This is the role played by BII.

Cottingham is clear that BII's involvement was a pivotal moment for his company. The initial funding and support from the Venture Lab program, which BII specifically designed to help scientific founders translate research into viable companies, was, he says, the "first measure of credibility" for the project in the eyes of the outside world. It was a signal that an organization like BII vetted the concept and deemed it worthy of support. But he is quick to emphasize that the true value of the partnership extends far beyond the financial.

As a non-profit, Nielsen explains, BII sits at the crossroads of academic science and commercial development, bringing together diverse players to strengthen the innovation pipeline. "Capital alone isn't enough," he says. "Building sustainable progress requires dedicated funding, experienced partners, and a collaborative community where founders can share lessons and navigate the challenges together."

Thanks to the BII's focus on women's health, Cottingham was connected with top-tier talent and a network of investors. This approach is by design. BII established a Women's Health Innovation Panel consisting of leading academics, industry players, investors, and other experts to assess applications and provide strategic support to startups. Furthermore, BII has fostered a community among its growing portfolio, which now includes 17 women's health startups, allowing for critical knowledge sharing. This commitment to nurturing the ecosystem is also reflected in the prestigious BII & *Science Translational Medicine* Prize for Innovations in Women's Health, which recognizes innovative advances with translational potential to impact women's health globally.

Despite the historical challenges, he remains hopeful about what lies ahead. "There's enormous potential waiting to be unlocked in women's health," Nielsen says, "And I believe we're only starting to see what's possible."

Sponsored by

BII BioInnovation Institute

RESEARCH INTEGRITY

Papers based on open health data face bans

PLOS, Frontiers, and others announce policies intended to stem the tide of suspect papers

CATHLEEN O'GRADY

Two major publishers have begun to automatically reject the vast majority of papers based on public health data sets, following revelations that unscrupulous actors use these data sets to churn out nonsense scientific papers.

Last month, PLOS and Frontiers both announced that submissions relying on data sets such as the U.S. Centers for Disease Control and Prevention–run National Health and Nutrition Examination Survey (NHANES), which collects diet and health data on more than 130,000 people, will not even be considered, unless the researchers do additional work to confirm their findings. Individual journals are imposing similar restrictions.

The new policies should help weed out fraudulent papers, at least in the short term, says Matt Spick, a statistician at the University of Surrey who brought the problems to light. But some researchers worry the rule could create unnecessary barriers for genuine research that can benefit public health. "It's good to have some guardrails," says Thu Nguyen, an epidemiologist at the University of Maryland. But "don't just throw the whole thing away and say it's all garbage."

Large public health data sets tend to contain a huge number of variables, making them fertile ground for studies that probe how various factors affect health. But their richness also allows unscrupulous actors to carve up the data in countless ways, seeking chance correlations that look like reliable statistical signals but are, in reality, just noise.

Emily Chenette, editor-in-chief of *PLOS ONE*, says PLOS journals have seen skyrocketing submissions of papers that find a statistical link between a health condition (such as depression) and a possible cause (such as vitamin D levels) in a specific group of people, such as women under age 35. Researchers have speculated that paper mill companies produce large volumes of papers like these—possibly written by artificial intelligence—and sell authorship slots to researchers.

> "Don't just throw the whole thing away and say it's all garbage."
>
> **Thu Nguyen**
> University of Maryland

Last month, the publisher announced it had "updated the standards against which we evaluate research using publicly available health and social science databases." Chenette says editors at PLOS journals will automatically reject papers based on these data sets unless the researchers have done extra legwork, such as experiments, to show their findings are robust. In the first month of the new policy, the rejection rate for such papers has increased from 40% to 94%, Chenette says.

Frontiers has announced a similar change, requiring editors to automatically reject papers that use public health data sets without some kind of "external validation." For instance, authors could use NHANES data alongside private data from their own institution, or run experiments to confirm their findings, says Elena Vicario, head of research integrity at Frontiers. Since the new policy was instituted in May, Vicario says, more than 5000 papers using open data sets have been rejected, including nearly all submissions using NHANES data.

Spick, whose work has found that Frontiers journals have published a significant chunk of problematic research, applauds the move. "They had let it get completely out of control," he says, "and now they have to take extra steps to restore credibility."

Some individual journals have also instituted defenses against the flood of junk papers. The *Journal of Global Health* announced in July that researchers using open data sets—including the Global Burden of Disease study and publicly available data from the UK Biobank, as well as NHANES—would now need to fill in an extensive checklist explaining what research had been conducted on their question before, and how their study contributes something new. Submissions of papers using open data sets have halved since then, says Igor Rudan, a global health researcher at the University of Edinburgh and editor-in-chief of the journal.

In an even more extreme move, the journal *Expert Opinion on Drug Policy* will no longer publish any

studies using the U.S. Food and Drug Administration Adverse Event Reporting System, an open data set tracking reports of medication side effects. The only exception to the journal's policy, which Retraction Watch reported last month, is research that has been invited by the journal's editors.

Scientific Reports has taken a different approach, relying on an internal policy that Spick, an associate editor at the journal, shared with *Science*. The policy alerts editors to the problems with NHANES and gives guidance on rejection. The journal's publishing director, Richard White, says Springer Nature has rejected thousands of NHANES papers across its journals, including at *Scientific Reports*. But the publisher considers training and guidance for editors to be more appropriate than a single blanket policy, White says.

Rumi Chunara, a biostatistician at New York University whose work sometimes uses public data, says the policies are reasonable and should result in better quality science. Checklists like the one introduced by the *Journal of Global Health* take a lot of time but "make for good-quality work," she says.

Spick suggests the organizations that create and maintain these data sets could address the problem by requiring researchers to submit their study design and rationale for review before gaining access to the data. But Chunara worries restricting access would make it harder for scientists working at institutions with few resources, such as those in the Global South, to conduct research. NHANES is a taxpayer-funded resource, she adds, so it should be publicly available.

Although the requirement for external validation may initially be effective, paper mills could easily sidestep it by looking for a spurious finding in two different data sets, Spick says. He thinks some organizations may already be employing this tactic, driving the rapid growth in papers that supplement findings from public health data with additional spurious correlations found in public genetic databases. Ultimately, no blanket policy will work properly if editors are asleep at the wheel, he says: "If editors are phoning it in, paper mill products will get through." □



Tens of thousands of fish were preserved during an early 20th century expedition by the *USS Albatross*.

PHOTO: JOHN WHALEN

MARINE SCIENCE

DNA from rum-soaked fishes chronicles century of change

Museum specimens reveal loss of genetic diversity in marine fishes of the Philippines **ERIK STOKSTAD**

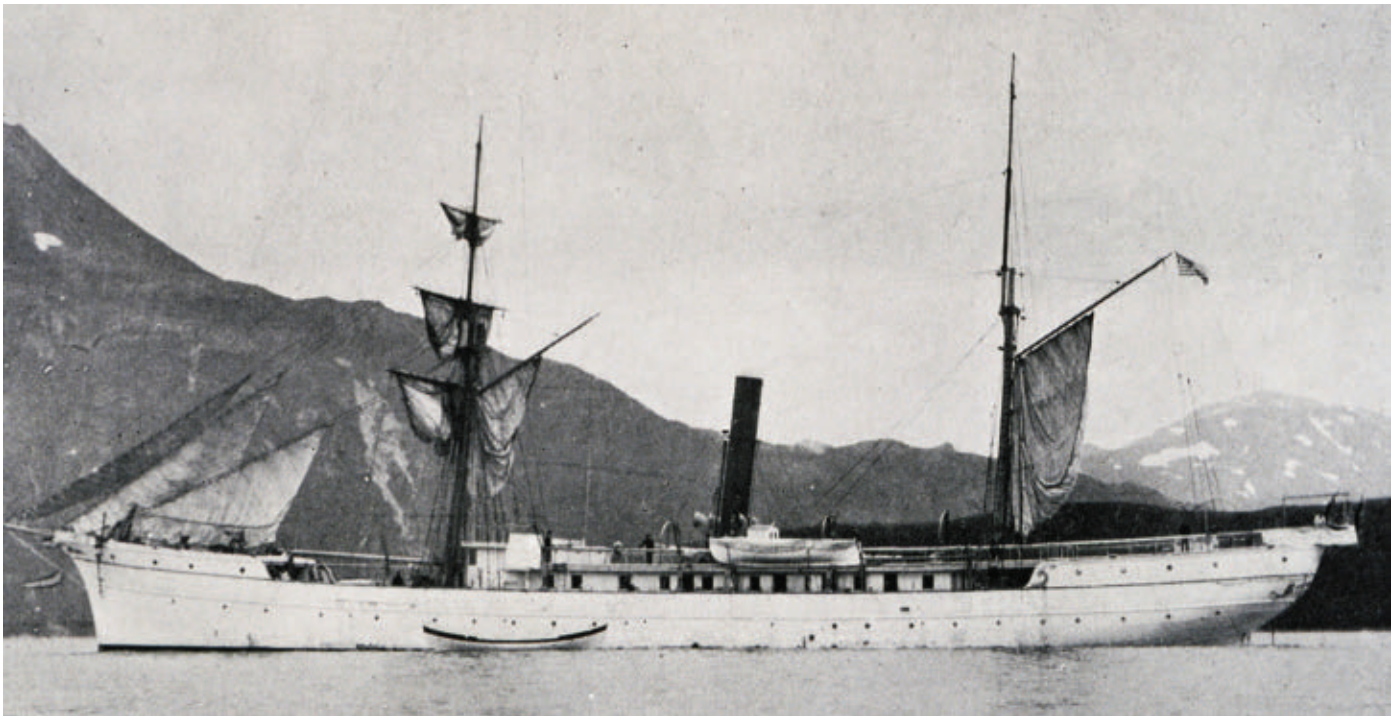
In October 1907, the *USS Albatross* departed San Francisco and steamed westward toward the Philippines. The 61-meter-long craft, operated by the U.S. Fish Commission, was the world's first ship purposely built for research, and this would be its longest expedition. For 3 years, researchers collected about 79,000 specimens of fish in the bays and reefs of the Philippines, documenting 117 new species.

Now, a team of ecologists and geneticists is probing those specimens to discover how the Philippines's rich marine biodiversity has changed after a century of increasing fishing and pollution. By comparing DNA in the historic specimens with that in modern fish populations, the team has discovered that as the fishes have grown scarcer, their genetic diversity has also declined, by 4% to 6% in three representative species. That's worrisome, says project co-leader Malin Pinsky of the University of California, Santa Cruz, because "genetic diversity provides the raw material for adaptation to novel conditions. Less genetic diversity means less ability to adapt to ongoing and future ocean change."

The results, which appeared last month on bioRxiv and in *The American Naturalist*, are only the first of a wide-ranging project to document changes in some of the world's most biodiverse fish communities. "This is really interesting stuff, because we don't know much about the history of such populations at all," says Bastiaan Star, an evolutionary biologist at the University of Oslo who wasn't involved.

About 10 years ago, project co-leader Kent Carpenter hatched the idea of turning to the DNA in the *USS Albatross* specimens to learn more about the region's fishes. The marine conservation biologist at Old Dominion University began to study marine life in the Philippines in the 1970s. Every time he's visited since then, the degradation of the seas has been more apparent. Some species have vanished entirely. "There have been dramatic changes," he says.

In 2017, Carpenter and an international team of researchers, including scientists at Silliman University and other institutions in the Philippines, launched a project to gather modern-day fish samples from the same areas the *USS Albatross* visited more than a century ago. Unlike the historic collectors, Carpenter's team didn't have a research vessel. Instead,



The *USS Albatross*, seen here in Alaska, was built in 1882 to carry out marine research.

PHOTO: NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION PHOTO LIBRARY

they largely purchased fresh fish from rural markets and harbors. So far, they have gathered more than 10,000 samples.

They also extracted DNA from species the *USS Albatross* collected in the greatest numbers. In a lucky break for modern-day geneticists, the crew had preserved the fishes using a cheap and plentiful liquid: high-proof rum. Most older museum specimens are preserved in formalin, which damages DNA. Alcohol, in contrast, leaves much more of the genetic material intact.

Another boon was the sheer size of the collection. Stored in the Smithsonian Institution's National Museum of Natural History, the *USS Albatross*'s haul sometimes included dozens of individuals of a single species, allowing the team to gauge how much genetic diversity was contained within the historic populations. Pinsky calls the project "the largest genome sequencing effort in the marine tropics."

One study, posted last month on bioRxiv, looked at two species of ponyfish (*Equulites laterofenestra* and *Gazza minuta*), fishes that live in shallow coastal waters and are intensively harvested. The work, part of René Clark's Ph.D. research at Rutgers University, found genetic diversity has declined by 6% in both species since the *USS Albatross* expedition.

To be sure the results didn't just stem from sequencing errors, which can appear more frequently in older, degraded samples, Clark says they tried "to kick this data every single way we could imagine to make sure it's real and not an artifact of it being historical DNA." They also found a hint of what caused the diversity loss: a severe population bottleneck. An analysis of the newer genomes indicated the number of reproducing ponyfish fell to just a few dozen sometime within the past 50 years or so, perhaps because of heavy fishing and pollution in its habitat.

Another study found a similar pattern for the blackbelted cardinalfish (*Taeniamia zosterophora*), a reef dweller. In a paper published last month in *The American Naturalist*, Kyra Fitz, a former Ph.D. student at Rutgers, and colleagues report its genetic diversity declined by 4% in the Cebu Strait—a location where fishing pressure is strong—but not at a more remote reef.

The declines exceed those typically measured for temperate species. In a 2014 meta-analysis of 140 economically valuable species such as cod, Pinsky found a 2% diversity loss, on average. The steeper declines in the tropical archipelago don't surprise Nina Overgaard Therkildsen, an evolutionary geneticist at Cornell University. Nearshore and reef fishes tend to have smaller populations to start with because their habitats are patchier and dispersal is often more constrained. And when population sizes get perilously low, rare alleles are more likely to be lost from the gene pool.

The studies did yield some good news. Fitz found that about 4% of genetic markers in the cardinalfish showed signs of being under selection, suggesting the fish are adapting to some of the changes around them. The evolutionary change is more pronounced in the Cebu Strait, where environmental impacts have been more intense. This evidence of selection is encouraging, says Steve Palumbi, a marine ecologist at Stanford University. "I think it's good news."

The team plans to add to its genetic findings by looking for physical changes the species have undergone since 1910. Team member Nadia Palomar-Abesamis, a marine ecologist at Silliman, is measuring both modern and historic specimens to look for changes in shape that may have resulted from habitat degradation and fishing pressure, such as shifts to faster, more streamlined bodies.

The genetic analysis also continues. By now the team has sequenced more than 5000 genomes across 39 species, hoping to find more clues to how the fishes have fared over the past century. "I'm definitely excited to see what comes next out of this," Therkildsen says. □

PALEOANTHROPOLOGY

Did lead poisoning help drive human evolution?

"Bold" hypothesis suggests tolerance for lead allowed *Homo sapiens* to outlive Neanderthals **MICHAEL PRICE**

It's one of the human story's most perplexing riddles: Why does our species, *Homo sapiens*, persist whereas close relatives with whom we once shared the planet, the Neanderthals and Denisovans, went extinct tens of thousands of years ago? One popular hypothesis suggests modern humans were better able to communicate and form bonds with one another, which helped them cope with climate swings and other existential pressures.

A study this week in *Science Advances* suggests how this social superiority may have emerged: Our brains were better protected against the toxic effects of lead. This scenario draws on evidence from fossil teeth that lead poisoning was rampant in our ancestors as long as 2 million years ago. The resulting brain damage might have created a ceiling on how socially complex hominins could be. Modern humans evolved defenses against lead, the idea goes, allowing them to build more complex and robust societies. Support for that idea comes from experiments with test tube minibrains—organoids made of brain tissue—engineered to have both modern and archaic versions of key brain genes.

"It's bold. It's a creative hypothesis," says Shara Bailey, a biological anthropologist at New York University who specializes in ancient teeth. "But there are a ton of limitations" to the study. For one thing, she notes, the teeth don't convincingly show ancient hominins were exposed to lead during early childhood, when it would have the greatest impact on brain development.

The hypothesis first occurred to Alysson Muotri, a neuroscientist and organoid expert at the University of California San Diego, when he attended a talk about lead exposure in Neanderthals by environmental epidemiologist Manish Arora of the Icahn School of Medicine at Mount Sinai. Arora later told Muotri his team had unpublished work showing exposure was widespread in ancient hominins. Lead is present in many natural environments, and human ancestors could have been exposed by drinking water that flowed through lead-bearing rocks or eating plants that absorbed lead from the soil.

At the time, Muotri's team was investigating the small suite of protein-coding genes, 61 in total, that differ between modern human genomes and those of Neanderthals and Denisovans. He was focused on one, *NOVA1*, that helps orchestrate neurodevelopment, turning on and off hundreds of other genetic switches in the brain.

In modern humans, *NOVA1* appears to favor complexity at the expense of slower maturation, whereas the variant in great apes—as well as in Neanderthals and Denisovans—has the opposite effect, Muotri explains. "A clear example is the fact that a baby chimp can outsmart a human babe, but the opposite is true in adult individuals." Muotri had been pondering how the modern *NOVA1* was selected. After talking to Arora, he wondered: Could lead have played a role? In modern humans, after all, childhood exposure to the neurotoxic element—for example through paint—has been linked to disordered emotional regulation and poorer impulse control and executive functioning.

Muotri and Arora decided to team up. They first looked for signals of lead exposure in 51 fossil teeth from ancient great apes, monkeys, and hominins found around the world. They found lead in 73% of their samples, including in ancient human ancestors such as *Australopithecus africanus*, the massive extinct ape *Gigantopithecus blacki*, ancient orangutans and baboons, Neanderthals, and early modern humans.

Next, Muotri and colleagues engineered brain organoids to have either the modern human version of *NOVA1* or the Neanderthal and Denisovan variant. They exposed the organoids to "very small, realistic amounts of lead that ancient humans might have encountered naturally," Muotri says.

Lead disrupted development in the two types of minibrains differently, they found. In the archaic organoids, it appeared to rattle neurons that express a gene called *FOXP2*. Mutations in *FOXP2* have been implicated in several communication disorders, suggesting it plays a key role in language. In the modern human organoids, these neurons remained essentially unharmed. "That was our eureka moment," Muotri says.

He says the results suggest that in modern humans, natural lead exposure selected for genetic variants, including *NOVA1*, that helped protect their brains from toxicity. That cleared the way for *H. sapiens* to become more socially cohesive and communicative—and outlive archaic cousins. "Sometimes evolution turns adversity into advantage," Muotri notes.

Bailey isn't sold. "At this stage, the evidence doesn't really convince me," she says. For one thing, "We don't know [at what age] these individuals were exposed to lead"—which is critical because children might have had different diets from adults. And strangely, the lead appeared to be concentrated in the ancient teeth's dentine, a calcified layer below the enamel, but not the enamel itself.



Lead may have dulled the social skills of ancient hominins and apes, including *Gigantopithecus blacki*, one of whose massive molars is shown here.

It's also not clear that exposing organoids to lead truly reflects what happens in the exposed brain during development, adds Takashi Namba, a neuroscientist at the University of Helsinki. Even so, the paper "is absolutely interesting," Namba says. "Not many studies have experimentally addressed environmental factors affecting human evolution." □

PHOTO: WANG WEI/XINHUA VIA ZUMA

Q&A

PHILIPPE AGHION

This year's Nobel Memorial Prize in Economic Sciences honors three researchers who have shown how innovation drives economic growth. Half of the prize, announced on 13 October, went to Joel Mokyr of Northwestern University for his discovery that people must understand the science underlying new innovations before they can lead to sustained growth. The other half is shared by Philippe Aghion of the College of France, INSEAD, and the London School of Economics and Peter Howitt of Brown University for showing that the ousting of established industries by newcomers—a process called "creative destruction"—is essential for growth. *Science* spoke to Aghion about his work. This interview has been edited for length and clarity.

Q: Your lab works on artificial intelligence (AI). How does creative destruction tie into this?
A: There is the fear with [AI] that job destruction would skyrocket. It's very important to understand that AI also has a productivity effect—firms that adopt AI become more productive. But you also need generous unemployment benefits and retraining. These institutional changes should accompany the AI revolution in order to minimize the negative effects and to magnify the positive.

Q: How can creative destruction help develop technology for fighting climate change?
A: Incumbent firms that innovated a lot in dirty technologies tend to continue spontaneously innovating in dirty technologies, while pressuring governments to prevent the entry of new firms and pro-green policies. New firms are not wedded to past bad habits, so creative destruction tends to push for green. If you allow for more creative destruction, you are helping green innovation.

Q: What do you make of the recent changes in economic and science policy in the U.S.?
A: I don't welcome them. When you get harsh on Harvard [University]; when you cut public research programs; when the government gets involved in what we should research and what we should not research and breaks the independence between research and government; when you restrict foreign students' mobility, all of this is not good news for an innovation-based economy.
—Cathleen O'Grady

NOBEL PRIZES

Architects of molecular cages win Chemistry Nobel

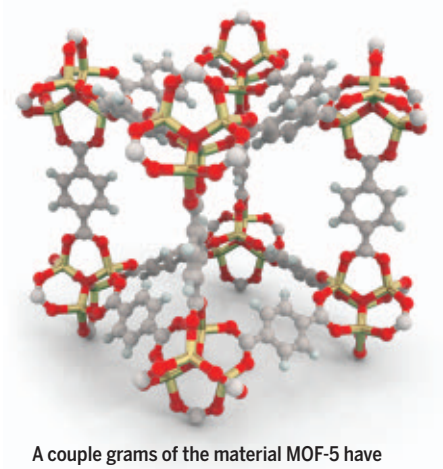
Spongelike materials called metal-organic frameworks can separate and store gases **DANIEL CLERY** and **CATHERINE OFFORD**

As early as the 18th century, people knew the pigment Prussian Blue had the mysterious ability to absorb and hold water. Its structure, deciphered much later, showed the compound was, in fact, a molecular cage: a cubic arrangement of iron ions joined by nitrile groups (carbon and nitrogen linked by a triple bond).

This year's Nobel Prize in Chemistry has gone to three researchers who pioneered modern versions of these porous structures by using organic molecules as struts to link up metal atoms. The Nobel-winning work, by Susumu Kitagawa of Kyoto University, Richard Robson of the University of Melbourne, and Omar Yaghi of the University of California (UC), Berkeley, led to metal-organic frameworks (MOFs) with large voids tailored to trap specific molecules, making them useful for storing gases such as hydrogen and methane, or soaking up water vapor or carbon dioxide out of the air. Some believe they even have the potential to deliver drugs or catalyze reactions.

"I'm delighted with this news," says Andrew Goodwin, a materials chemist at the University of Oxford. "A lot of people have been waiting for it. It really developed a new field in chemistry, and it's very clearly having a major international impact."

The initial breakthrough came from Robson in the 1980s. He created a tetrahedral molecule using nitrile groups to link up copper ions at each of the four vertices. Moreover, he was able to stack these molecules into a repeating crystalline structure with large cavities. "It was just amazing imagination that he had to devise, in fairly simple terms, this way of making his new class of materials," says Brendan Abrahams, a longtime collaborator of Robson's at Melbourne.



A couple grams of the material MOF-5 have the surface area of a football field.

IMAGE: RAMON ANDRADE 3DCIENCIA/SCIENCE PHOTO LIBRARY

WINNERS

SUSUMU KITAGAWA, Kyoto University

RICHARD ROBSON, University of Melbourne

OMAR YAGHI, University of California, Berkeley

Robson's material was quite fragile and prone to collapse. But Kitagawa saw the potential. In the 1990s, despite meager funding, he made a thin 2D structure that could hold acetone molecules in its cavities. In 1997, he and his team succeeded in making a 3D crystal with cobalt, nickel, or zinc ions as the hubs. It was threaded with open channels, enabling it to absorb and release methane, nitrogen, and oxygen. And it was stable. "At first, everyone thought such structures would fall apart immediately—that was common sense," Kitagawa said last week at a press conference organized by Kyoto University. "But we proved that stable, robust frameworks could, in fact, be built this way.

In 1999, Yaghi, then at the University of Illinois Urbana-Champaign, revealed the work that would cement the status of metal-organic frameworks. Instead of linking up ions, Yaghi and his team used more complex metal clusters as the hubs. Their iconic material, known as MOF-5, was a cubic lattice with huge cavities. A couple grams of the material have the surface area of a football field.

The researchers also showed they could tailor the voids and their properties. By using longer or shorter or-

ganic linkers and exposing metal ions in the pore spaces, they could tune the materials to attract specific gases at mild pressures and release them as the pressure drops. The number of design possibilities is immense, Yaghi said last week at a press conference organized by UC Berkeley. "If you think it, you can make it."

Today, researchers have conjured up tens of thousands of kinds of metal-organic frameworks, and dozens of companies are trying to commercialize their use. They have "huge importance for humanity," says Wendy Lee Queen, an inorganic chemist at the Swiss Federal Institute of Technology, Lausanne.

Some companies are trying to use them to store methane fuels for cars, whereas others see potential for trapping hydrogen, a clean fuel that otherwise requires high pressures, cold temperatures, or large volumes to store.

Another application is to use the selectivity of the materials as a filter to soak up particular compounds. Yaghi's group has spun off companies developing metal-organic frameworks to pull pure water out of thin air—even in desert environments. Other companies are testing the materials to extract carbon dioxide from industrial processes and power plants to reduce greenhouse gas emissions.

The materials could also be used to sequester toxic wastes—or simply to identify when they are present. Jing Li, an inorganic and materials chemist at Rutgers University, has experimented with using metal-organic frameworks to detect explosives and toxic vapors. Now, she is developing some that can separate mixes of hydrocarbons into various petrochemical products. Metal-organic frameworks could be more efficient than traditional approaches such as distillation, she says. "There has been tremendous progress."

In theory, metal-organic frameworks could even be used as catalysts: Reagents could be implanted in the cavities, and reactions could progress through different steps as compounds pass through the corridors of the material. At the Kyoto press conference, Kitagawa said his dream was to use the materials not just to separate and capture the components of air—nitrogen, oxygen, water vapor, and carbon dioxide—but to convert them into useful products. "From these simple elements—C, H, O, N—we make proteins, food, and fuels," he said. "Air, in that sense, is 'invisible gold.'" □

With reporting by Cathleen O'Grady, Dennis Normile, and Robert F. Service.

IN FOCUS



PHOTO: YVA MOMATIUK AND JOHN EASTCOTT/MINDEN

A weevil takes wing

An adult rice weevil (*Sitophilus oryzae*) sits on top of a grain of rice with wings spread, as if ready to fly away and find a mate, in this magnified view by Zhang You of Kunming, China. The image of the tiny insect was announced this week as the winner of the annual Nikon Small World competition, which attracts photomicrography submissions from scientists and artists from around the world. Rice weevils are a major pest of rice, wheat, and corn worldwide. Female insects gnaw into the grains and lay eggs inside; their larvae feed on the grain, pupate, and ultimately emerge as adults. Weevils with their wings spread are hard to find, You notes in a press release, so it was by sheer chance that he discovered one preserved on a windowsill. For a sense of scale, he mounted it on a rice grain before photographing it. —Matthew Warren



Congress's failure to pass 2026 spending bills has forced a government shutdown.

PHOTO: E. BILLMAN/*SCIENCE*

TRUMP ADMINISTRATION

Chaos and confusion as U.S. shutdown drags on

Mass layoffs, and subsequent reversals, have added to research spending woes ***SCIENCE* NEWS STAFF**

"It feels like getting hit by a car right after you've barely survived a train wreck." That's how one U.S. government scientist describes having her agricultural research facility largely shut down because of a political standoff over funding even as it struggles to cope with hefty staff reductions and potentially deep budget cuts. "It's just been one blow after another."

The battering isn't likely to end soon, as Republicans and Democrats continue to spar over how to end the U.S. government shutdown that began on 1 October—and President Donald Trump's administration has launched a chaotic, and possibly illegal, effort to fire thousands of federal employees.

Last week, for example, the administration sent layoff notices to at least 1300 health workers, including key disease specialists at the Centers for Disease Control and Prevention (CDC), only to rescind roughly half of the notices within hours. Meanwhile, thousands of other government scientists remain locked out of their offices, laboratories, and email accounts, putting research projects and grantmaking activities in limbo. Many others are still reporting to work, but without pay, because they are in roles deemed "essential"—such as operating spacecraft, treating patients in clinical trials, or keeping laboratory animals alive.

The shutdown and firings come on the heels of 8 months of turmoil that has already seen the Trump administration fire or push into retirement numerous staff at research agencies, freeze hiring, and restrict purchasing and travel. Many scientists are reeling from grant cuts—some later reversed by court orders. Science agencies had to scramble to distribute money by the end of the 2025 fiscal year on 30 September after Trump policies delayed grantmaking early in the year. The White House has also proposed deep cuts to research spending in the 2026 fiscal year that have left many researchers unable to plan for the future.

"It's slowdown on top of slowdown," says Benjamin Feldman, who runs a core facility that makes transgenic zebrafish for experiments at the National Institutes of Health's (NIH's) child health institute.

Feldman has been going into his facility a few times a week for "time sensitive things" such as testing fish to ensure they are bred at the optimal age. But otherwise NIH "is pretty quiet," Feldman says, speaking for himself and not NIH.

About 25% of NIH employees are considered essential, mostly within NIH's in-house research program, but most extramural grants staff are furloughed, meaning grant-review meetings are on hold. Reviewers can continue to upload their critiques, however, although NIH can't process new applications.

The disruption couldn't come at a worse time for Julie Kovacs, a bioinorganic chemist at the University of Washington (UW) whose grant proposal was slated to be reviewed on 23 October. Any delay, she says, could

leave her research group—which investigates how certain enzymes fight tumors and synthesize antibiotics—in limbo next spring, when their current NIH grant ends.

Kovacs and other researchers would typically rely on bridge funding from their institution to stay afloat. But UW has warned its employees that "current federal policy volatility" may affect such funding. Kovacs fears she might be forced to shut her 30-year-old lab: "Such a waste," she says.

At CDC and its parent agency, the Department of Health and Human Services (HHS), staffers are suffering from whiplash. First came sweeping layoffs unveiled on the night of 10 October. Those fired included leaders of efforts to confront a measles outbreak, scientists with CDC's Epidemic Intelligence Service, and nearly the entire staff of the *Morbidity and Mortality Weekly Report*, the agency's flagship journal. Within 24 hours, however, HHS officials rescinded many of those layoffs, saying about 700 were errors. As *Science* went to press, other senior CDC officials, including the head of its chronic disease prevention center, were waiting to hear whether their firings would also be reversed.

The American Federation of Government Employees (AFGE), which has 820,000 members, has sued to block the firings. "The administration is misusing the shutdown process for partisan ends and violating the very laws that govern how shutdowns are supposed to function," says Everett Kelley, AFGE's president.

Furloughed federal scientists aren't supposed do work-related tasks at home. Several, from NASA and other agencies, told *Science* they are continuing to work on manuscripts, sometimes using secure, third-party apps to communicate with colleagues. But they won't submit papers to a journal while the government is closed, for fear of punishment.

At the U.S. Department of Agriculture (USDA), the shutdown has only added to concerns about the future of the many research programs and facilities the Trump administration has proposed eliminating or downsizing. A wave of firings and resignations has already hobbled some programs, say two USDA researchers. And a vague administration threat to deny backpay to furloughed workers has only "reinforced incentives for the departure of early or midcareer employees," one says.

At the National Science Foundation, the shutdown has complicated life for NSF-funded scientists needing an icebreaker to work in Antarctica. In 2013, a 16-day shutdown disrupted U.S. Antarctic research because the company providing logistical support ran out of money. NSF fixed that problem by giving its contractors sufficient advance funding to maintain operations at the three U.S. Antarctic stations during future shutdowns. But earlier this year NSF decided to cancel the lease on the *RV Nathaniel B. Palmer*, the only U.S. research vessel dedicated to serving the region.

The agency has pledged to find substitutes for the several upcoming cruises scheduled for the *Palmer*. But making those switches requires NSF, the contractor, and academic groups to work out myriad logistical details. The shutdown, however, has furloughed nearly all NSF staff, hindering those interactions.

Such disruptions are likely to continue well into October, as neither Democrats nor Republicans appear ready to call a truce in their fight over 2026 funding. One thing is certain, however: The longer the shutdown lasts, the more pain researchers will feel. □

With reporting by Phie Jacobs, Jocelyn Kaiser, David Malakoff, Jeffrey Mervis, and Hannah Richter.

PHOTO: HYOUNG CHANG/*THE DENVER POST*

IN FOCUS

'Magic' mushrooms evolved twice independently

Psychoactive compounds in certain mushrooms can provoke profound psychedelic experiences—but researchers don't fully understand why the fungi evolved. A new analysis has added to the confusion. Two distinct genera of magic mushroom produce the same hallucinogen, psilocybin, through entirely different chemical pathways, according to a 21 September paper in *Angewandte Chemie International Edition*. The researchers made the discovery after examining the genes and proteins involved in psilocybin production in a fiber cap mushroom (*Inocybe corydalina*). They assumed these mushrooms would synthesize the compound in the same way as the better studied *Psilocybe* fungi, such as *P. cubensis* (pictured). But to their surprise, the process was completely different, suggesting it evolved independently in the two mushrooms. Evolution might have favored mushrooms that produce psilocybin because it exerts a discombobulating effect that deters insects from munching on them—or because it encourages animals that enjoy the trippy experience to pass on their spores.
—Taylor Mitchell



COLD STORAGE

On an Arctic archipelago, frozen soil may preserve a hidden history of viruses

KAI KUPFERSCHMIDT, on Svalbard
Photography by **LUKASZ LARSSON WARZECHA**



Sébastien Calvignac-Spencer stores one of the soil samples he and his team collected over 3 days on the Arctic archipelago of Svalbard in July.

As you drive from the airport into Longyearbyen, the largest town on the Norwegian archipelago of Svalbard, a concrete structure juts from a sandstone mountain on the right. It's the entrance to the Svalbard Global Seed Vault, where more than 1 million samples of wheat, rice, maize, and other crops are stored underground as a hedge against global disaster.

The vault is located here because even if the cooling system that keeps the seeds refrigerated at –18°C fails, the frozen ground should ensure the seeds' preservation for hundreds of years. Here, about 1300 kilometers from the North Pole, the earth itself functions as a freezer.

That's also the reason Sébastien Calvignac-Spencer came to visit Svalbard in mid-July. An evolutionary biologist at the Helmholtz Institute for One Health, Calvignac-Spencer is interested in the evolution of viruses whose genomes are made of RNA—a more fragile molecule than its cousin DNA. Surviving thousands of years, ancient DNA sequences have changed scientists' understanding of the history of bacterial diseases such as the plague and syphilis, and of smallpox, caused by a virus with a DNA genome. They have also revolutionized our understanding of human origins and migrations.

Ancient RNA could hold secrets of its own, including how the viruses that cause influenza, polio, and Ebola have evolved and how their relationships to animals and humans have changed with, for example, the industrial revolution or the rise of modern agriculture. But until now, very little ancient RNA has been sequenced. Few published examples are older than 100 years—and some of them are widely believed to be lab errors. "There's almost nothing there," says Edward Holmes, an expert on viral evolution at the University of Sydney.

Calvignac-Spencer acknowledges that RNA generally degrades more rapidly than DNA. But he believes its fragility has become a dogma that keeps researchers from even looking for it. He has had a string of successes in recent years recovering old genomes of measles, influenza, and other viruses from formalin-fixed tissue samples in museums. On Svalbard, he's hoping to show that in the cold, dry permafrost, RNA viruses' genomes can survive much longer—perhaps thousands or even hundreds of thousands of years. "They have a seed vault here, maybe they have a virus vault, too," he says.

Hendrik Poinar, who heads the Ancient DNA Centre at McMaster University, is eager to see what Calvignac-Spencer comes up with. He agrees RNA's fragility has sometimes been overstated and predicts new techniques, some of them being developed in his own lab, may soon give the field a huge boost. "I think there'll be a revolution in this space very shortly," Poinar says.





LONGYEARBYEN, the northernmost town in the world, is an unusual place. In the summer, the Sun does not set for about 4 months and the population of some 2000 doubles with an influx of tourists and researchers. The constant light can be both energizing and a little disorienting. Carrying a rifle is mandatory outside the city limits because of the risk of a polar bear attack. (The last fatality was a Dutch visitor killed at a campsite close to the airport in 2020.) Calvignac-Spencer wasn't looking forward to the bear risk when he planned the trip. "I'm not a field person, so I was like hmmm," he says. He stopped reading about it.

Having spent his career in offices and laboratories, "I was sort of like surprised myself to be in Svalbard," Calvignac-Spencer says. In other ways, however, he had ended up exactly where he wanted to be.

As a child growing up in Toulouse, France, Calvignac-Spencer was fascinated by dinosaurs and dreamed of being a paleontologist. For his Ph.D., he joined an ancient DNA lab at the University of Lyon, working partly on brown bears and partly on retroviruses in humans and apes. "This made me realize I'm actually interested in the evolution of infectious agents," he says. In 2010, after completing his Ph.D., he joined Fabian Leendertz, a wildlife veterinarian then at the Robert Koch Institute who was studying pathogens in chimpanzees in Taï National Park in Ivory Coast.

The work in Berlin did not involve ancient specimens, but Calvignac-Spencer closely followed the field of ancient DNA, which was exploding at the time. In 2017, when their group moved into a new building in Berlin,

To collect pieces of permafrost from deep in the soil, Tim Urich (right) and other team members dig a hole and then hammer a hollow metal tube into the soil.

CREDITS: (PHOTO) LUKASZ LARSSON WARZECHA; (MAP) V. PENNE/*SCIENCE*



On a plastic plate, Lilli Gramma (background) and Sébastien Calvignac-Spencer cut tiny pieces out of an earth core.

Leendertz and Calvignac-Spencer decided to turn some spare laboratory space into a paleogenetics lab, partly to extend their work on the disease ecology of great apes into the past.

For Calvignac-Spencer, it was also an opportunity to start looking at the evolution of infections more generally. DNA researchers have studied pathogens from thousands of years ago, but he saw a niche in the analysis of RNA viruses from more recent specimens. For one thing, he says, the past 200 years have seen momentous changes that could have influenced viral evolution—think of the industrial revolution, massive changes in farming, or the widespread adoption of vaccines.

Researchers can infer part of the history of a virus by comparing modern genomes and using the rate of genomic change to calculate when their common ancestor existed, and what it may have looked like. But because RNA viruses mutate fast, this strategy does not reach very far back in time. The last known common ancestor of the respiratory syncytial virus, for example, stems from the 20th century; six of the eight segments of the flu genome trace back to the 19th century. "So we quickly come to a sort of a phylogenetic horizon, and we don't see beyond that," Calvignac-Spencer says.

Ancient RNA can bypass that limit, but finding it is hard. The bacteria that cause syphilis and leprosy leave marks on human skeletons that can help researchers find them. RNA viruses rarely leave any trace, "so it is a case of actually processing a lot of material, hoping you get lucky," says Lucy van Dorp, a microbial geneticist at University College London. Calvignac-Spencer decided to try pathology collections. He started with an old collection housed close to his lab, at the Charité Museum of Medical History Berlin. For a first trial run, he chose 12 formalin-fixed specimens from the early 20th century.

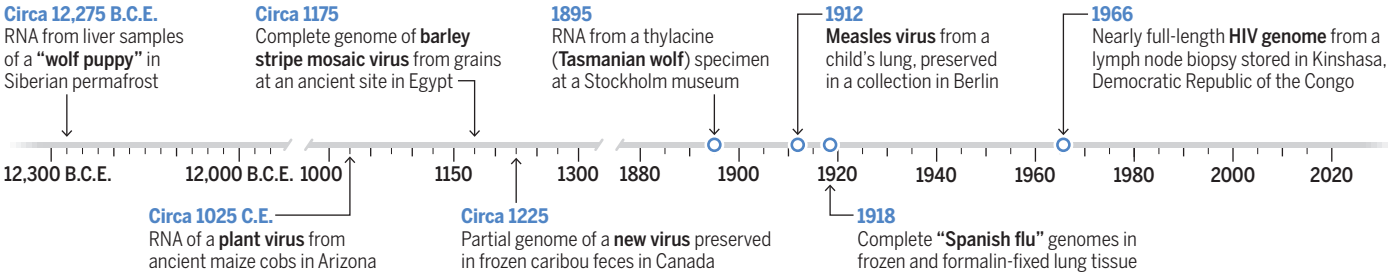
One was the lung of a 2-year-old child who had died of measles in 1912. Calvignac-Spencer managed to piece together an entire viral genome from it, still the oldest human RNA virus to surrender its genetic blueprint. In a paper published in *Science* in 2020, he used this genome and more modern ones to pinpoint when measles might have jumped from cattle to humans: not in the Middle Ages, as earlier research suggested, but as early as 500 B.C.E., just as cities were growing big enough to sustain continuous circulation.

Other samples from that first trawl through the collection yielded one complete and two partial sequences of the virus that caused the 1918 influenza pandemic, also known as the Spanish flu. A team led by virologist Jeffery Taubenberger, now acting director of the U.S. National Institute of Allergy and Infectious Diseases, had first sequenced that virus in 2005, largely from samples

PHOTO: LUKASZ LARSSON WARZECHA

Stepping back in time

Ancient RNA has proved much harder to find and sequence than ancient DNA, which goes back hundreds of thousands of years. Most RNA studies have concentrated on well-preserved specimens in medical collections from the 19th and 20th centuries; a few have salvaged RNA from samples that are hundreds or thousands of years old.



taken from a patient buried in permafrost in Alaska. (That genome helped explain why the virus was so deadly, especially in young adults, unlike other flu strains.)

With four hits out of 12 samples, "I thought, wow, OK, that's a 30% success rate in a first batch," Calvignac-Spencer says. "Then we went all in." His team has sampled human and animal specimens from collections in Vienna; Rome; Bologna, Italy; Paris; Madrid; Munich; and many other cities, which it is now analyzing. His success has inspired others, Van Dorp says. "Whenever I talk to him, he's always very confident," she says. "He's like, 'Well, you have to try,' and he's had quite a good success rate doing so."

Yet few pathological specimens go back further than the 19th century. Calvignac-Spencer's oldest sample so far is from a 1836 rabies patient, stored in a collection in Bologna, that he has yet to analyze. "So then the question was: What are the other archives we could think of?" he says. One answer was Svalbard. Countless generations of nesting seabirds and other wildlife had likely shed viruses that might linger in the permafrost there. But sampling it sounded easier than it turned out to be.

ON THE AFTERNOON of the first day of the expedition, Calvignac-Spencer was standing on a grassy slope a few minutes outside of Longyearbyen, surveying a square hole he and his colleagues had just dug. The researchers had hit stones and gravel about a half-meter down, and the thin layer of soil above did not look promising. "It's completely thawed," said team member Tim Urich, a soil microbiologist at the University of Greifswald. "Our task will be to find places where the soil part is deeper." As huskies barked and howled in the distance, the researchers started to look for the next spot to dig.

Finding Svalbard soil that has remained frozen for hundreds or thousands of years was just one of many challenges, but it was a crucial one. Only there did ancient RNA stand a chance of surviving.

There are several reasons for RNA's fragility. The sugars in its backbone have an extra oxygen that makes it more likely to be cleaved. And RNA is usually single stranded, whereas DNA typically consists of two strands, entwined in a double helix that adds stability. But the biggest problem are ribonucleases, small enzymes that destroy RNA. Organisms use them to metabolize superfluous cellular RNAs and as a defense against pathogens, including RNA viruses. Ribonucleases, also known as RNAses, are found everywhere and are notoriously stable, Poinar says. "That is the biggest thing you are fighting."

Researchers once thought RNA, which is far more abundant in cells than DNA, might persist through sheer quantity. In 1985, the year Svante Pääbo, then a Ph.D. student at Uppsala University, published a seminal paper on DNA extracted from mummies, Franco Rollo at the University of Camerino described fragments of RNA found in a 1600-year-old cress seed. In a letter to *Nature* a few years later titled "Mummy RNA lasts longer," Rollo argued that RNA likely made up a majority of all nucleic acids in ancient samples.

But ancient DNA quickly won the day as a research topic, and researchers have used it to study life further and further into the past. In 2013, they published the genome of a horse that lived more than half a million years ago. In 2022, they sequenced DNA of numerous plants and animals frozen in Greenlandic ice for more than 2 million years, providing a glimpse of the ecosystem back then. RNA fell by the roadside, as ancient DNA drew most paleo-



Samples, stored in small plastic tubes and logged, are stored in a dewar, essentially a thermos flask cooled with liquid nitrogen.

genomics funding. By now, Poinar says, extraction and purification methods for ancient nucleic sequences "are geared towards DNA, they're not geared towards RNA."

Yet Rollo may have had a point, says Tom Gilbert, an evolutionary biologist at the University of Copenhagen. In the early 2000s, he and Mike Worobey of the University of Arizona studied decades-old formalin-fixed samples stored in the Democratic Republic of the Congo, looking for HIV. "What was striking to me was that we were often finding better RNA preservation than DNA," Gilbert says. "And I'm like,

CREDITS: (GRAPHIC) A. MASTIN/*SCIENCE*; (PHOTO) LUKASZ LARSSON WARZECHA

that doesn't make any sense, RNA is really fragile."

But that truism might not hold, he notes, if ribonucleases are deactivated—which both formalin and cold can do. Once they're gone, researchers would be just as likely to find RNA from 2 million years ago as DNA, Gilbert says: "I don't really think there is a limit." But Holmes isn't persuaded. "I just can't see that [RNA] will provide the same perspective on the past as ancient DNA. How can it?" he says. "But that's not to say that we won't get some ancient RNA viruses from some really well preserved cases."



Frozen penguin mummies from Antarctica, on the opposite pole from Svalbard, could also harbor old RNA viruses, Sébastien Calvignac-Spencer thinks.

The COVID-19 pandemic has stoked the interest in ancient RNA because it could hold clues to the evolution of coronaviruses related to SARS-CoV-2, Van Dorp says. "RNA viruses are a hot topic. They're superinteresting." One of the world's most prominent coronavirologists, Christian Drosten, recently released a preprint with the sequence of an RNA virus isolated from a bat collected in northern Germany in 1919 and stored in ethanol in Berlin's Natural History Museum. It was a bandavirus not known to infect humans, but similarly old samples of coronaviruses could help pin down where and when OC43 and 229E, cousins of SARS-CoV-2 that cause the common cold, first started to infect humans, Drosten says.

A natural RNA virus archive on Svalbard would open the way to other studies as well, for example on the influenza evolution before the 1918 pandemic. Calvignac-Spencer wasn't quite sure what the team might find; dozens of viruses might be hiding in the cold soil. For now, he just wanted to prove the archive exists.

THE REST OF THE first day, the group hauled its gear across the Arctic tundra, stopping here and there to dig a hole. But nowhere was the ground frozen, even 1 meter down. "It's sort of a practical personal experience of climate change," Calvignac-Spencer said. Indeed, Svalbard has seen a steeper temperature rise than most other places, and scientists have to dig deeper each summer to hit the permafrost. The virus vault, if it exists, may be breaking down—another reason to do the research now, Calvignac-Spencer said. "We want to see if this is worth pursuing, also with the view that this may be disappearing."

On the second day, the team went to a series of steep cliffs where thousands of birds, especially little auks, congregate during breeding season. Reindeer grazed close by and an arctic fox wandered around. "The idea is to sample in these places where huge amounts of animals have been living sometimes for centuries and millennia," Calvignac-Spencer said. The researchers dug again, and here and there they hit what seemed to be frozen ground. A team member drove a long, hollow metal cylinder into the ground with a hammer, pulling out a cylindrical core of dirt.

Wearing masks, gloves, and plastic sleeves to keep modern viruses from contaminating the samples, Calvignac-Spencer and Ph.D. student Lilli Gralla knelt on the grassy ground and cut tiny chunks out of the cores. Gralla dropped them in tubes and deposited them in the dewar, essentially a huge thermos flask cooled by liquid nitrogen.

THE SVALBARD EXPEDITION isn't the first to push the search for ancient RNA beyond museum specimens. In 2014, researchers published a partial genome of a new virus found in frozen caribou feces in Canada, dated to about 700 years ago. That same year, researchers reported a complete genome of a virus isolated from 750-year-old barley grains in Egypt. In 2015, another paper described a plant virus sequenced from ancient maize cobs found in Arizona (see timeline, p. 233).

Calvignac-Spencer is also trying his luck in Antarctica. In 2023, he read a French newspaper article about the work of Steve Emslie, an ornithologist and archaeologist at the University of North Carolina Wilmington who essentially studies penguin history. Since 2001, Emslie has been going to Antarctica regularly, looking for old breeding sites of Adélie penguins. When he finds them, he digs into the ice—and thus the past—layer by layer, finding eggshell fragments, feathers, bones, and even entire mummified birds, sometimes thousands of years old.

Calvignac-Spencer immediately wondered whether the mummies contained ancient RNA viruses. He contacted Emslie, who sent him samples from the mummies' throat, trachea, ribs, and cloaca. Calvignac-Spencer says his lab has been able to piece together the genomes of several viruses, including a near-complete rotavirus dated to about 1900 years ago.

Like other claims of ancient RNA, the work, not yet published, is bound to face questions about lab contamination. A 1999 study reporting tomato mosaic virus in ice that was up to 140,000 years old

PHOTO: STEVEN EMSLIE



The cemetery near Longyearbyen holds burials from the 1918 influenza pandemic. A 1998 expedition to find the virus failed because bodies were not well preserved.

PHOTO: LUKASZ LARSSON WARZECHA

faced skepticism, for example. "I do not believe this paper because these ancient samples are so similar to modern ones," Holmes says. Because RNA viruses mutate so fast, he says, "If you see a sequence of an ancient or a supposedly ancient RNA virus and it's identical to a modern virus, you know something is not right."

Calvignac-Spencer admits the penguin rotavirus lacks two typical signs of being ancient. It does not show the distinctive U-shaped distribution of damage at the ends of its RNA fragments, known as a "smiley pattern"—although that may just be because of the exceptional preservation of the penguin, he says. And when he added the sequence to a family tree of related viruses, it did not sit on a short branch, as expected of a virus that is different from modern viruses and closer to an ancestral virus. But Calvignac-Spencer says the long stretches of RNA may have survived because of the exceptional preservation of the penguin. He also notes that rotaviruses have an uncommon, double-stranded RNA genome, which may be more durable.

The penguin mummies are the best-case scenario for RNA preservation, Calvignac-Spencer says: "They have been exposed to very cold conditions and very dry conditions. You cannot get any better than that."

He hopes to find the same on Svalbard. On the third day of the expedition, Calvignac-Spencer and the other team members went to another site, where colleagues had spotted some exposed permafrost. In the cold wind, they dug, cored, and sampled again, much of the work routine by now. The next day they boxed up their samples and flew back to Germany.

ON A HILLSIDE overlooking Longyearbyen lies an old cemetery, just a few rows of white crosses defying the eternal wind. It marks a previous effort to delve into viral history on Svalbard.

In 1998, Canadian medical geographer Kirsty Duncan led an expedition here to exhume the bodies of seven miners who had died of the Spanish flu in 1918. The coffins were not as deep in the ground as expected, possibly because of frost heave—a phenomenon that led Svalbard authorities to outlaw coffin burials in 1950. The bodies were not well preserved. And unlike Taubenberger's group, which took its samples in Alaska the year before, Duncan's team did not manage to isolate any fragments of the influenza virus.

Calvignac-Spencer's expedition, 27 years later, may come up empty, too. He puts the chance of finding any ancient virus in his samples, now safely stored in a Berlin freezer, at maybe 10%. "But at this point, before I say something is not possible, I will just try it," he says—and try again if needed. He's already planning another virus-hunting trip to Svalbard. □

PERSPECTIVES

CELL BIOLOGY

Kiss, shrink, run

A unified mechanism directs synaptic vesicle release

Katharina Lichter

Synaptic vesicles are tiny membrane spheres that store and transport neurotransmitters. They enable communication between neurons in the brain by fusing with the presynaptic membrane to release neurotransmitters into the synaptic cleft, a process known as exocytosis. This mechanism is indispensable for synaptic transmission. Yet the precise mode of vesicle fusion remains unresolved, and there are competing theories about whether synaptic vesicles fully collapse or fuse only transiently and about how rapidly they are recycled. Resolving these discrepancies is challenging because these processes occur within milliseconds at nanometer dimensions. On page 258 of this issue, Tao *et al.* (*1*) report using time-resolved cryo–electron tomography to visualize synaptic vesicle dynamics in situ. They propose a unified model in which vesicles transiently fuse and release neurotransmitters through a stable pore before downsizing and detaching intact.

More than 70 years ago, electron micrographs provided the first detailed view of the synaptic complex: presynaptic and postsynaptic neurons separated by a narrow cleft, with a cluster of vesicles each around 40 nm in diameter on the presynaptic side (*2*). To Bernard Katz, the neurophysiologist and later Nobel Prize winner, these vesicles offered a structural correlate for his kinetic theory that neurotransmitters are released not continuously but rather in discrete packets, or quanta (*3*). More than a decade later, the concept of active zones—specialized presynaptic compartments regulating timely vesicle release—was introduced (*4*). By the 1970s, it was evident that vesicle exocytosis alone could not sustain transmission at these sites; vesicles also had to be retrieved, recycled, and refilled to maintain the presynaptic vesicle pool (*5*, *6*). This progression of discoveries set the stage for a central debate: What precisely occurs at the moment of vesicle fusion? One view holds that after initial fusion, vesicles can fully collapse into the presynaptic membrane. This would then require complete reformation of the vesicle through membrane invagination (endocytosis) to maintain the presynaptic vesicle pool. Both the location and timescale of this synaptic vesicle endocytosis vary. It can take place inside and outside of the active zone, and clathrin-independent ultrafast endocytosis can happen within 50 to 100 ms after an action potential (*7*), whereas clathrin-dependent modes can occur seconds later. A parallel hypothesis is that vesicles remain largely intact, fusing with the presynaptic membrane transiently to release their cargo before pinching off again (*6*). This process, called "kiss-and-run," bypasses the need for full endocytosis.

Tao *et al.* tackled the challenge of resolving synaptic vesicle fusion events at millisecond temporal and nanometer spatial resolutions. Using cultured rat hippocampal neurons, they combined optogenetic stimulation, ultrarapid freezing, and three-dimensional cryo–electron tomography to capture vesicles in distinct states before, during, and after fusion at excitatory synapses. Each method is essential: Optogenetic stimulation ensures selective activation of defined neurons, ultrarapid freezing preserves vesicle states that last only nano- to milliseconds, and cryo–electron tomography reconstructs vesicles as full spheres rather than two-dimensional cross sections, enabling accurate volume quantification without fixation-induced alterations. Using these methods, the authors report a release mode in which vesicles first "kiss" (semifuse) with the presynaptic membrane and then discharge neurotransmitters through a stable pore and "shrink" to a surprisingly constant fraction of their original volume within a few milliseconds. Although some vesicles then fully collapse into the presynaptic membrane, most of them "run" away (detach) from it as intact organelles by around 70 ms after the initial action potential—a hyperfast form of vesicle recycling. According to Tao *et al.*, 80% of release events may proceed through this so-called kiss-shrink-run pathway, suggesting that it could represent the dominant form of synaptic vesicle exocytosis. The findings link two pathways once thought distinct—full collapse of the vesicle with endocytosis and kiss-and-run—into a unifying sequence (see the figure).

How does this release mode functionally connect to the vesicle states that precede fusion? Synaptic vesicles progress through a sequence of steps, consisting of tethering, docking, and priming, to prepare for fusion. Recent work suggests that docking is a tran-

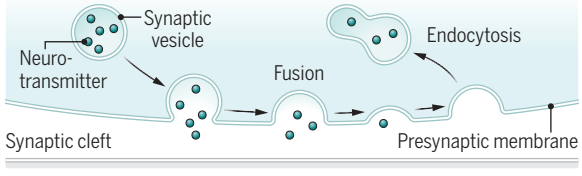
Two models combine

The arrival of an action potential (AP) at a presynaptic terminal causes synaptic vesicles to fuse to the presynaptic membrane, releasing neurotransmitters into the synaptic cleft. The kiss-shrink-run model of this fusion process shares elements of two previously proposed models.



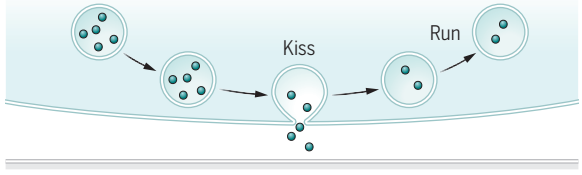
Full-fusion model

After fusion, vesicles collapse and must be reformed through endocytosis.



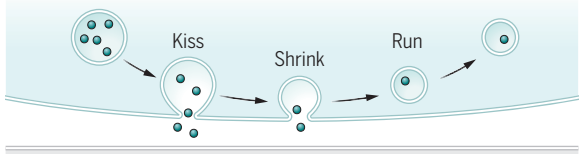
Kiss-and-run model

Vesicles fuse transiently to release neurotransmitter before detaching.



Kiss-shrink-run model

Vesicles fuse transiently and release neurotransmitter through a stable pore before shrinking and detaching.



GRAPHIC: A. MASTIN/*SCIENCE*

sient process, forming a calcium-dependent equilibrium with undocking (*8*), which may underlie delayed release and activity-dependent changes in synaptic strength (plasticity). How this equilibrium biases the initiation and the selection of the reported release modes, kiss-shrink-run and kiss-shrink-collapse, remains unresolved. Building a continuous framework for the entire vesicle cycle will require probing these intermediate states in a similar number of synapses as was sampled by Tao *et al.* (~1000), but over a longer train of action potentials.

Tao *et al.* focused their analyses on excitatory synapses, and it is not yet clear whether the "kiss-shrink-run" model can be generalized to all synapse types, especially inhibitory synapses. Network stability depends on a balance between excitation and inhibition, with inhibitory neurons providing rapid control over circuit activity. Indeed, vesicles at γ-aminobutyric acid–releasing inhibitory synapses have distinct functional properties (*9*). To understand how excitatory-inhibitory balance within neural circuits is maintained with the necessary millisecond precision (*10*), future work must establish whether excitatory and inhibitory synapses rely on common release mechanisms. Another key question is whether all classes of synaptic vesicles rely on the kiss-shrink-run process. For instance, at hippocampal mossy fiber synapses, which are critical for memory storage, vesicles vary in size, and specialized populations such as dense-core vesicles may modulate synaptic strength and timing (*11*, *12*). Tao *et al.* highlight membrane tension and curvature as relevant factors for vesicle fusion, raising the possibility that these factors act selectively on different vesicle types. Such selectivity could in turn shape synaptic output by modulating transmitter load and release probability.

How does the kiss-shrink-run release pathway contribute to synaptic plasticity? Presynaptic vesicle pools have been proposed to encode short-term memory as "pool engrams" (*13*, *14*). Changes in vesicle number and volume as well as spatial arrangements (*12*, *14*) could all serve as storage elements, extending vesicles from passive carriers to active substrates of computation at single synapses. Tao *et al.* show that the recycled vesicles do not immediately reintegrate into the reserve pool, raising the question of how kiss-shrink-run interacts with vesicle maintenance: Under what conditions does a vesicle persist in an engram-like state, when does it transition into release, and how is a recycled vesicle recruited back into an engram? These functional states are likely to shape both the dynamics of plasticity and the speed of vesicle recycling, with direct implications for memory (*15*). □

REFERENCES AND NOTES

1. C.-L. Tao *et al.*, *Science* 390, eads7954 (2025).
2. S. L. Palay, G. E. Palade, *Anat. Rec.* 118, 335 (1954).
3. P. Fatt, B. Katz, *J. Physiol.* 115, 320 (1951).
4. R. Couteaux, M. Pécot-Dechavassine, *C. R. Acad. Hebd. Seances Acad. Sci. D* 271, 2346 (1970).
5. J. E. Heuser, T. S. Reese, *J. Cell Biol.* 57, 315 (1973).
6. B. Ceccarelli, W. P. Hurlbut, A. Mauro, *J. Cell Biol.* 57, 499 (1973).
7. S. Watanabe *et al.*, *Nature* 504, 242 (2013).
8. G. F. Kusick *et al.*, *Nat. Neurosci.* 23, 1329 (2020).
9. C. Park *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 118, e2022133118 (2021).
10. S. Loomba *et al.*, *Science* 377, eabo0924 (2022).
11. D. Vandael, P. Jonas, *Science* 383, eadg6757 (2024).
12. K. Lichter *et al.*, *Cell Rep.* 40, 111382 (2022).
13. S. A. Josselyn, S. Tonegawa, *Science* 367, eaaw4325 (2020).
14. D. Vandael, C. Borges-Merjane, X. Zhang, P. Jonas, *Neuron* 107, 509 (2020).
15. K. M. Harris, *J. Physiol.* 10.1113/JP286983 (2025).

ACKNOWLEDGMENTS

The author thanks P. Jonas for feedback on the manuscript and acknowledges support from the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie grant agreement no. 101034413.

10.1126/science.aec0091

Institute of Science and Technology Austria (ISTA), Klosterneuburg, Austria.
Email: katharina.lichter@ist.ac.at

HYDROGEN PRODUCTION

Making hydrogen production durable

An interphase shields polymer electrolytes from electrochemical oxidation in an alkaline environment

Yu Seung Kim

Hydrogen is widely viewed as a low-carbon energy system, serving both as a clean fuel and as a medium to store intermittent renewable power. An electrolyzer—a device that electrochemically splits water into hydrogen and oxygen—is an essential component for hydrogen production. Anion-exchange membrane electrolyzers combine the benefits of the traditional liquid alkaline system with those of more advanced proton-exchange membrane electrolyzers, by using nanoparticle catalysts (*1*). Although this technology can eliminate the need for corrosive liquid electrolytes and costly precious metal catalysts (*2*), its durability under pure-water operation is far behind those of precedents. On page 294 of this issue, Hou *et al.* (*3*) report a strategy that substantially improves durability of an anion-exchange membrane electrolyzer by incorporating an inorganic molecular additive to shield the polymer electrolyte from electrochemical oxidative degradation. This points to a new pathway for stabilizing anion-exchange membrane electrolyzers, strengthening their potential for cost-effective hydrogen production.

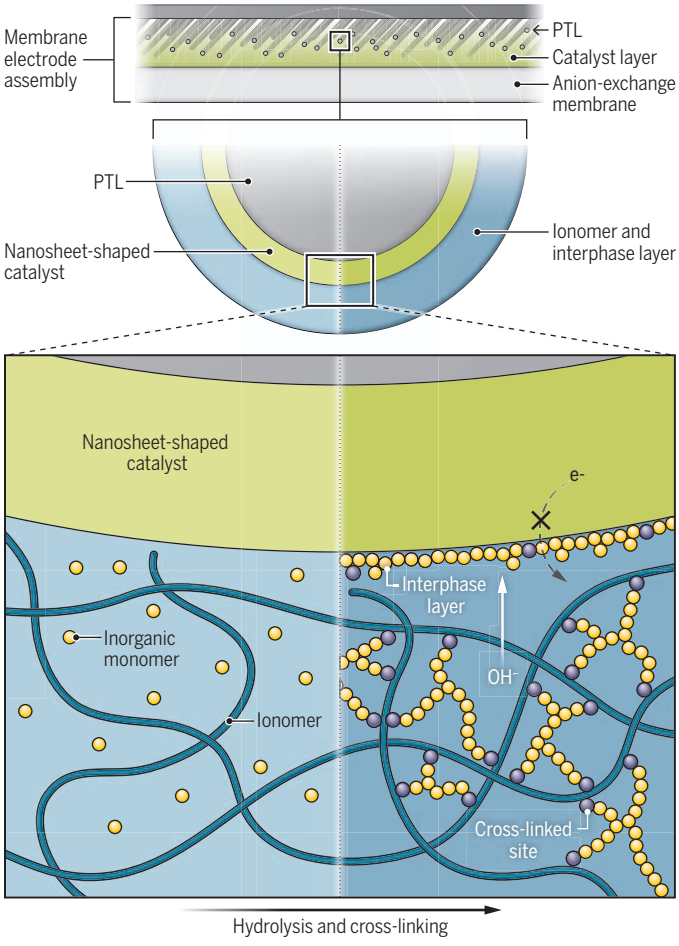
This points to a new pathway for stabilizing anion-exchange membrane electrolyzers... for cost-effective hydrogen production.

Electrochemical reactions proceed more rapidly in strong alkaline environments (pH > 12), permitting the use of inexpensive catalysts such as nickel, iron, silver, and carbon. This reduces reliance on precious metal catalysts such as platinum or iridium, which contribute to the cost of hydrogen production (*4*, *5*). However, ionomers (polymers that contain ionic groups) used as polymer electrolytes in the electrodes of anion-exchange membrane electrolyzers are vulnerable to degradation under device operating conditions. Specifically, ionomers experience chemical breakdown under high-pH conditions, which dictate durability of the devices into which they are incorporated (*6*). In addition, adsorption of polymer fragments on catalyst surfaces, coupled with electrochemical oxidation of ionomers, has been shown to undermine performance (*7*, *8*). Consequently, achieving both durability and efficiency in alkaline electrolysis using an anion-exchange membrane remains unresolved (*9*).

Inspired by the interphase design of lithium-ion batteries in which organic and inorganic components of the solid electrolyte passivate electrodes against electrolyte decomposition (*10*), Hou *et al.* applied this concept to anion-

"Interphasing" the boundary

Inorganic monomers, such as zirconium propoxide, can be polymerized to interact with ionomers (polymers with ionic groups). This process creates cross-linked networks that form an organic-inorganic interphase between the catalyst and the ionomer, which protects the ionomer from electrochemical oxidation in an alkaline environment.



e-, electron; OH^-, hydroxide ion; PTL, porous transport layer.

exchange membrane electrolyzers. The authors used metal-alkoxide (zirconium propoxide) additives that covalently bonded to the ionomer through polymerization. This created a layer of organic-inorganic interphase (zirconium oxo and hydroxo oligomer) between the ionomer and the anode catalysts (see the figure). The interphase exhibited structural motifs that protected the ionomers from electrochemical degradation. Double-bonded oxygen and hydroxyl (-OH) groups coordinated to zirconium resisted oxidation and impeded electron transport from the catalyst to the ionomer. Carboxylic (-COOH) end groups interacted with quaternary ammonium moieties of the ionomer, reinforcing ionic conductivity while slowing degradation. The enhanced dispersion of charged catalyst nanoparticles formed a chemically and mechanically stable interphase. The oligomeric framework of the interphase prevented additives from leaching under hydrated conditions, while retaining sufficient moiety to assemble into a coherent interphase. Collectively, these features produced a protective layer that improves hydroxide conductivity across the ionomer-catalyst boundary, thereby boosting the electolyzer's catalytic performance.

Beyond stabilizing the ionomers at the anode, the engineered interphase also mitigated the migration of oxidative species during water splitting, protecting the anion-exchange membrane from secondary degradation. The pure-water alkaline anion-exchange membrane electrolyzers created by Hou *et al.* demonstrated degradation rates below 0.49 mV $hour^{-1}$ under industrially relevant conditions (2.0 A cm^{-2} and 70°C), which is a >20-fold improvement compared to conventional systems without the interphase layer.

The work by Hou *et al.* could be broadly applied to polymer-based electrochemical devices such as fuel cells and co-electrolyzers (devices that convert water and carbon dioxide to value-added products), where operation at high electrode potentials (>0.6 V) is unavoidable. Polymer electrolytes offer important advantages in these systems by eliminating corrosive liquid electrolytes and salt precipitation during operation. However, these advantages come at the cost of vulnerability: Exposure to high electrode potentials drives electrochemical oxidation of polymer electrolytes and generates degradation products that compromise both materials stability and the local reaction environment. Although ionomer-free architectures (*11*) have achieved good electrochemical stability under high electrical potentials, using this strategy often means losing the benefits that polymer electrolytes offer, such as stabilization of highly dispersed nanoparticles, effective water management, and robust adhesion to the membrane.

Hou *et al.* demonstrated interphase stabilization of a phenylene group–rich commercial polymer electrolyte (PiperION). Perfluorinated polymers such as Nafion have long dominated electrochemical devices because of their oxidative stability, but they tend to degrade under alkaline conditions. With increasing scrutiny of perfluoroalkyl substances owing to environmental and health concerns (*12*), the strategy of Hou *et al.* not only extends the durability of ionomers in alkaline electrolyzers but also may be transferable to acidic environments. This could help displace perfluoroalkyl-based materials in electrochemistry systems (*13*). Robust, phenylene-based electrolytes could therefore offer a pathway toward environmentally friendly and sustainable devices.

Important questions remain. Durability of the interphase layer over thousands of hours of operation under industrially relevant conditions needs to be examined. In addition, its performance under dynamic load cycling (charging and discharging at varying electric potentials) and at high current densities has not been tested. But if the life spans of pure-water electrolyzers become comparable to or surpass those of conventional electrolyzers that use potassium hydroxide as a feedstock (*14*, *15*), it could unlock the full potential of pure-water anion-exchange membrane electrolyzers for scalable hydrogen production. □

REFERENCES AND NOTES

1. D. Li *et al.*, *Nat. Energy* 5, 378 (2020).
2. M. Muhyuddin *et al.*, *Chem. Rev.* 125, 6906 (2025).
3. S. Hou *et al.*, *Science* 390, 294 (2025).
4. K. Yue *et al.*, *Science* 388, 430 (2025).
5. X. Yang *et al.*, *Adv. Energy Mater.* 14, 2400029 (2024).
6. D. Li *et al.*, *Energy Environ. Sci.* 14, 3393 (2021).
7. I. Matanovic, Y. S. Kim, *Curr. Opin. Electrochem.* 38, 101218 (2023).
8. G. A. Lindquist *et al.*, *Energy Environ. Sci.* 16, 4373 (2023).
9. J. Lim *et al.*, *ACS Energy Lett.* 9, 3074 (2024).
10. Y. Gao *et al.*, *Nat. Mater.* 18, 384 (2019).
11. S. Y. Kang *et al.*, *ACS Appl. Energy Mater.* 8, 2567 (2025).
12. E. J. Park *et al.*, *Chem. Soc. Rev.* 53, 5704 (2024).
13. J. H. Choi *et al.*, *ACS Energy Lett.* 10, 2392 (2025).
14. L. Yin *et al.*, *Angew. Chem. Int. Ed.* 64, e202503715 (2025).
15. C. Hu *et al.*, *Angew. Chem. Int. Ed.* 63, e202316697 (2024).

ACKNOWLEDGMENTS

The author used an artificial intelligence–assisted tool (ChatGPT-5) to improve the readability of this Perspective. The author acknowledges support from the US Department of Energy, Office of Energy Efficiency and Renewable Energy, and Hydrogen and Fuel Cell Technologies Office through the Hydrogen from Next-Generation Electrolyzers of Water (H2NEW) Consortium.

10.1126/science.aeb7886

Materials Synthesis and Integrated Devices (MPA-11), Los Alamos National Laboratory, Los Alamos, NM, USA. Email: yskim@lanl.gov

GRAPHIC: N. BURGESS/*SCIENCE*

NEUROSCIENCE

Neural compass in the sky

Head-direction neurons maintain stable directional signals during large-scale navigation in the wild

Yue-Qing Zhou[1,2] and **James J. Knierim**[1,2,3,4]

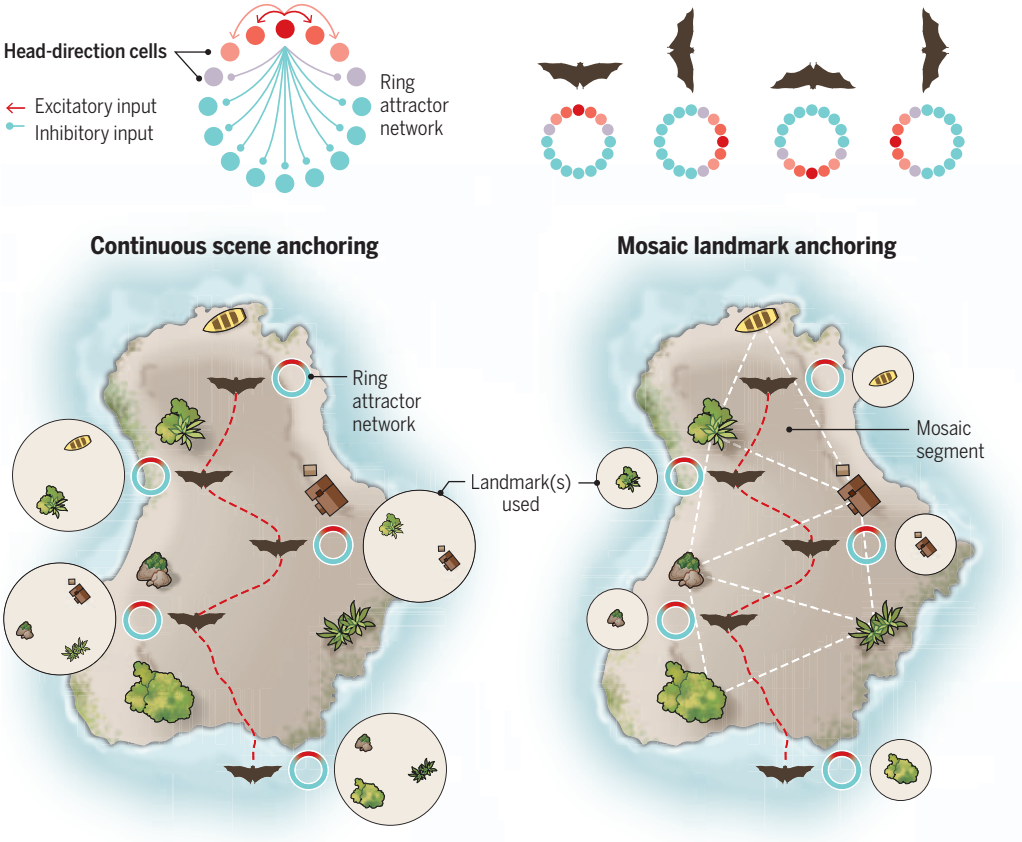
Head-direction cells are neurons in the brain that increase their firing rates when animals face specific directions. The head-direction network serves as an internal compass, converting sensory cues into an instantaneous estimate of head direction to support navigation. Decades of laboratory and theoretical work have generated a sophisticated understanding of the mammalian head-direction network, including how external cues and internal self-motion cues contribute to precise head-direction coding. Yet most evidence comes from animals confined to small, laboratory environments, typically ~1 to 2 m^2 in area. On page 259 of this issue, Palgi *et al.* (*1*) present recordings of the activity of head-direction cells from Egyptian fruit bats made as the bats flew over large areas on a small, remote island. The data show that preferred firing directions of head-direction cells remain stable and anchored to a global reference frame in this large, natural environment just as they do in the laboratory.

Head-direction cells were initially identified in freely moving rats in the dorsal presubiculum, a brain area involved in spatial navigation located in the parahippocampal region (*2*, *3*). These neurons increase their activity when the rat's head points in a specific direction in an allocentric reference frame, regardless of the animal's location or the head's orientation relative to the body. On the basis of those rodent studies, the almost universally accepted model of the generation and maintenance of head-direction representations is a theoretical "ring attractor," in which head-direction cells maintain a self-sustained "bump" of activity on a recurrently connected, hypothetical ring of cells (*4*–*6*). Input from vestibular and other self-motion cues shift this bump around the ring in concert with head movements by integrating angular head velocity, a process that inevitably accumulates noise. Stable external landmarks provide anchors to minimize and correct these errors, thus keeping the bump aligned to the external world (*7*). In a demonstration of presumed convergent evolution of a computational algorithm, a physical ring-like attractor network of putative head-direction cells was discovered in the fruit fly (*Drosophila melanogaster*) in a part of the brain called the central complex, providing strong support for the theoretical models derived from the rodent studies (*8*). The ring attractor thus offers a powerful framework to understand the interplay between internal head-direction representations and external cues.

To determine whether this framework applies outside of the artificial laboratory environment, Palgi *et al.* developed a miniature wireless data logger to simultaneously record single-neuron activity and centimeter-resolution location information from Egyptian fruit bats. The bats were released to explore a 350- by 250-m island off the coast of Africa, and the activity of head-direction neurons was recorded over several days. Recording electrodes were positioned in two brain regions known to contain head-direction cells, the presubiculum and retrosplenial cortex, a medial cortical area that bridges visual and hippocampal circuits. Although cells that encode self-location information (such as place cell and grid cells) and head-direction cells are traditionally considered as spatial representations that anchor to a global frame of reference, the results of more recent studies (*9*, *10*) suggest that the self-location representations of large environments may be composed of multiple spatial fragments rather than forming a single, continuous map. These results raise a funda-

Anchoring of head-direction representations

Head-direction cells maintain a self-sustained "bump" of activity (red) on a theoretical ring of cells called a ring attractor network. Input from self-motion cues shifts this activity around the ring as head direction changes. As a bat flies over a large island with many landmarks, the preferred firing directions of head-direction cells remain consistent. Thus, the same cells in the ring attractor network are active when the bat is facing north, regardless of the animal's location on the island. However, the mechanism underlying this visual anchoring remains unknown. Two models are proposed. In the continuous scene anchoring model, the head-direction network continuously associates its firing with the scene visible at each location. In the mosaic landmark anchoring model, a small set of reliable landmarks are selected to stabilize the head-direction network while the bat is in a specific segment of the island.



GRAPHIC: K. HOLOSKI/*SCIENCE*

mental question: Does the head-direction compass remain globally consistent in its directional tuning in a large-scale, outdoor environment, or do preferred firing directions of head-direction cells also form a mosaic, with each head-direction cell having a different preferred direction at different parts of the island? Palgi *et al*. showed clearly that the head-direction cells in the presubiculum and retrosplenial cortex maintain parallel preferred firing directions everywhere across the ~87,500 m^2 area of the island. The preferred directions changed noticeably during the bats' first night of exploration of the island and became stable after three to six nights of experience. The results suggested that head-direction cells outdoors become anchored to a global frame over time, similar to the findings of head-direction cells recorded in the laboratory.



The behavior of head-direction cells in the Egyptian fruit bat (*Rousettus aegyptiacus*) is usually studied in animals confined to small, laboratory environments, rather than in the natural world.

How is such robust directional coding achieved over such a large area? The head-direction ring attractor model posits that during initial exploration, neural representations of visual landmarks that are perceived when the animal's head faces a certain direction become associated with the head-direction neurons that represent that direction, thus assuring that the attractor activity bump stays aligned to the external world. In the small spaces of the laboratory environment, the associations are straightforward because even a single landmark that is always (or almost always) visible can serve as the anchor point for the entire head-direction network across the entire environment [assuming that the problem of visual parallax is solved—the differences in egocentric views of an environment when the rat's head is in a particular direction but the rat's body is at different locations (*11*)]. However, in the real world with a much larger scale of space and complex sets of landmarks such as trees, mountains, and shorelines, the binding of the attractor bump to the world becomes more complicated because the visual cues that could serve as landmarks are not consistently visible from all locations. On the island, for example, the visual panorama the bat perceived facing north at the southern shore shared little resemblance to the view facing north at the northern shore. One solution might be that the bats used celestial cues (the Moon and stars) as anchoring landmarks, similar to the use of the Sun compass in some species (*12*). Palgi *et al*. demonstrated that this explanation is unlikely because the Moon's changing position in the sky as a function of the lunar cycle would require a more complex neural ephemeris table (a table of astronomical positions over time used by navigators) than that of the Sun, and the head-direction cells maintained stable directional tuning even when cloud cover masked the stars. A geomagnetic compass also seems implausible because there is no evidence that bats possess a geomagnetic sense. The system must therefore use terrestrial landmarks.

How might the binding of terrestrial landmarks to the head-direction ring attractor be accomplished over such large areas? One possibility is that the ring attractor network continually associates representations of constantly changing visual scenes (visual backgrounds) with specific head directions (*13*). Such a strategy will demand new computational models to uncover the learning rules that support the continuous association of visual assemblies to the ring attractor network in the endlessly changing visual input that the bat perceives as it flies through the world. Another possibility that would reduce computational complexity would be that bats split the global cognitive map into fragments (*9*), and within each fragment, the head-direction network chooses a subset of specific, reliably visible landmarks as directional inputs to stabilize the whole head-direction ring attractor network. In this view, the compass remains globally consistent, but its specific landmark anchors are chosen in a mosaic fashion (see the figure). This strategy will require a computational model to understand the process of weighing different cues to serve as directional landmarks, how these landmarks are accurately incorporated into the head-direction network, and how the "hand-off" between different anchoring landmarks is accomplished as the bat crosses the boundaries of different segments. Testing these models will require a return to the laboratory, where landmark cues can be manipulated in controlled ways, which can then lead to further predictions about how the system would behave back in the field.

Thus, outdoor recordings offer the indispensable "ground truth" for the navigation network in nature, whereas well-controlled laboratory experiments can dissect the underlying mechanism of the formation of cognitive representations. Together, these experiments can drive computational models that are mechanistically precise and ecologically valid, and the resulting principles should guide the design of the next generation of artificial navigation algorithms, which are necessary for the self-localization and mapping functions in autonomously navigating robots, self-driving cars, and other applications (*14*, *15*). □

REFERENCES AND NOTES

1. S. Palgi *et al*., *Science* 390, eadw6202 (2025).
2. J. S. Taube, R. U. Muller, J. B. Ranck Jr., *J. Neurosci.* 10, 420 (1990).
3. J. S. Taube, R. U. Muller, J. B. Ranck Jr., *J. Neurosci.* 10, 436 (1990).
4. K. Zhang, *J. Neurosci.* 16, 2112 (1996).
5. W. E. Skaggs, J. J. Knierim, H. S. Kudrimoti, B. L. McNaughton, *Adv. Neural Inf. Process. Syst.* 7, 173 (1995).
6. K. S. Kakaria, B. L. de Bivort, *Front. Behav. Neurosci.* 11, 8 (2017).
7. B. L. McNaughton, L. L. Chen, E. J. Markus, *J. Cogn. Neurosci.* 3, 190 (1991).
8. J. D. Seelig, V. Jayaraman, *Nature* 521, 186 (2015).
9. K. J. Jeffery, *Learn. Behav.* 52, 19 (2024).
10. D. Derdikman *et al*., *Nat. Neurosci.* 12, 1325 (2009).
11. A. Bicanski, N. Burgess, *J. Neurosci.* 36, 11601 (2016).
12. A. Rozhok, *Orientation and Navigation in Vertebrates* (Springer, ed. 1, 2008).
13. F. Steinbeck *et al*., *Adapt. Behav.* 32, 407 (2024).
14. T. C. Knowles, A. G. Summerton, J. G. H. Whiting, M. J. Pearson, *Biomimetics (Basel)* 8, 399 (2023).
15. D. Ball *et al*., *Auton. Robots* 34, 149 (2013).

10.1126/science.aeb6967

[1]Zanvyl Krieger Mind/Brain Institute, Johns Hopkins University, Baltimore, MD, USA. [2]Kavli Neuroscience Discovery Institute, Johns Hopkins University, Baltimore, MD, USA. [3]Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD, USA. [4]Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, USA. Email: jknierim@jhu.edu

PHOTO: STEVE GETTLEVY/MINDEN PICTURES

PHYSICS

Quantum trions in equilibrium

Quantum fluid of trions is demonstrated in two-dimensional material systems

Tianyi Ouyang and **Su-Fei Shi**

Excitons—neutral quasiparticles of a negatively charged electron bound to a positively charged hole—are ubiquitous in semiconductors and optoelectronic devices. Incorporating an extra charge carrier (positive or negative) into an exciton produces additional charge and spin degrees of freedom, converting excitons into fermions with increased effective mass. Joining three fermions creates a trion, which often requires optical excitation. However, these quasiparticles are short-lived, which limits the study of their effects on materials used in optoelectronic devices or their application in quantum optics (*1*). On pages 299 and 304 of this issue, Qi *et al.* (*2*) and Nguyen *et al.* (*3*), respectively, report experimental evidence of trions in a state of balance, in which these quasiparticles are formed and then recombined at an equal rate, extending their existence. This state of equilibrium may allow further exploration of trion behavior and new quantum states that are composed of many particles.

Two-dimensional (2D) materials, such as transition metal dichalcogenides, are suited for exploring many-body bound states in which three or more particles interact strongly and behave as a single entity (*4*). For example, heterostructures held together by weak van der Waals forces, such as twisted, stacked layers of two different 2D crystalline materials, can be engineered into strongly coupled layers in which 2D planes of electrons and holes are spatially separated by a thin sheet of insulating material. This construction enables independent electrostatic control of charge carrier density in each layer, allowing for systematic tuning of the total charge density and the electron-hole balance (neutral charge condition). Additionally, the recombination of electrons and holes is suppressed because of the large momentum mismatch between them. This results in a long-lived "fluid" of interlayer excitons in a state of thermal equilibrium—a state in which the excitons' interactions dominate their individual, particle-like behavior (*5*–*8*). Although these stable excitons have been observed in gallium arsenide–based quantum well structures and bilayer graphene systems (*9*–*11*), the small effective mass of excitons and the relatively large interlayer distance have prevented the use of existing materials to form higher-order particle complexes, such as trions. Because the binding energy of trions is typically an order of magnitude smaller than that of excitons, trions require strong attraction (Coulomb coupling) between electrons and holes along with a reduced kinetic energy of the charge carriers.

Qi *et al.* and Nguyen *et al.* experimentally demonstrated that thermal equilibrium fluids of interlayer trions can form in strongly coupled monolayers of transition metal dichalcogenide by reducing the thickness of a boron nitride sheet between them. In the study of Qi *et al.*, the electron density in molybdenum disulfide and the hole density in tungsten disulfide were independently controlled by applying an interlayer bias voltage (an electric potential difference between the layers of 2D materials). The authors observed absorption resonances that emerged precisely when the carrier densities matched electron-to-hole ratios of 2:1 or 1:2. This result confirmed the formation of positive and negative interlayer trions. Notably, two identical charge carriers in a trion formed a spin-singlet state in which two electrons or two holes exhibited paired, antiparallel spins. Qi *et al.* estimated the binding energy of a trion to be about 1 meV. Further doping the transition metal dichalcogenides with impurities resulted in a double-layer system with several distinct phases, such as electron and hole Fermi gases, an interlayer exciton fluid, positive or negative trion fluids, and mixtures of different quasiparticle states. Laser excitation revealed additional optical spectra, which were interpreted as higher-order many-particle bound states, such as four-particle (tetron) and five-particle (penton) complexes. These states appeared only transiently after photoexcitation and were not equilibrium phases. Nonetheless, they indicate that the platform of Qi *et al.* can host more-complex many-body bound states.

Nguyen *et al.* investigated the same bilayers with a thicker hexagonal boron nitride sheet (1.5 to 2 nm compared with less than 1 nm in Qi *et al.*). When the hole density became twice the electron density, the authors detected the formation of a trion fluid with a binding energy of 1 meV. The trion fluid displayed insulating behavior at a low density of charge carriers. However, as the density increased, the fluid underwent a continuous crossover into an electron-hole plasma (a high-density mixture of free electrons and holes), giving rise to a spin-singlet state. Applying an out-of-plane magnetic field of several tesla dissociated the trions. The study of Nguyen *et al.* establishes trions as long-lived equilibrium quasiparticles, which confirms theoretical predictions that electron-hole bilayers can host trion fluids and other higher-order states.

...the equilibrium quantum trion fluid sets the stage for many exciting possibilities.

The ability to produce equilibrium interlayer trion fluids could drive the formation of new correlated states, such as Wigner crystals (*12*). A trion's heavy effective mass can suppress its kinetic energy, forcing trions into a specific periodic lattice structure. In addition, trion fluids provide a platform for studying correlated states of Bose-Fermi mixtures in which excitons, trions, and free carriers coexist. A rare structure in which two trions pair with each other in a cloud of excitons may also be possible (*13*). Whether a trion pair could give rise to an exotic state of matter, such as the Bose-Einstein condensation or Bardeen-Cooper-Schrieffer superconducting regime, remains to be seen (*13*, *14*). Furthermore, a correlated mixture of equilibrium trions and excitons could enable unconventional superconducting pairing, leading to topological superconductivity (*15*). The experimental realization of the equilibrium quantum trion fluid sets the stage for many exciting possibilities. □

REFERENCES AND NOTES

1. H. Wang *et al.*, *Phys. Rev. B* **93**, 045407 (2016).
2. R. Qi *et al.*, *Science* **390**, 299 (2025).
3. P. X. Nguyen *et al.*, *Science* **390**, 304 (2025).
4. X. Chen, Z. Lian, Y. Meng, L. Ma, S.-F. Shi, *Nat. Commun.* **14**, 8233 (2023).
5. L. Ma *et al.*, *Nature* **598**, 585 (2021).
6. R. Qi *et al.*, *Nat. Commun.* **14**, 8264 (2023).
7. P. X. Nguyen *et al.*, *Science* **388**, 274 (2025).
8. R. Qi *et al.*, *Science* **388**, 278 (2025).
9. E. Tutuc, M. Shayegan, D. A. Huse, *Phys. Rev. Lett.* **93**, 036802 (2004).
10. X. Liu, K. Watanabe, T. Taniguchi, B. I. Halperin, P. Kim, *Nat. Phys.* **13**, 746 (2017).
11. J. I. A. Li, T. Taniguchi, K. Watanabe, J. Hone, C. R. Dean, *Nat. Phys.* **13**, 751 (2017).
12. T. Smoleński *et al.*, *Nature* **595**, 53 (2021).
13. J. von Milczewski, X. Chen, A. Imamoglu, R. Schmidt, *Phys. Rev. Lett.* **133**, 226903 (2024).
14. M. Y. Kagan, A. Bianconi, *Condens. Matter* **4**, 51 (2019).
15. C. Zerba, C. Kuhlenkamp, A. Imamoğlu, M. Knap, *Phys. Rev. Lett.* **133**, 056902 (2024).

ACKNOWLEDGMENTS

The authors acknowledge support from the National Science Foundation (DMR-1945420, DMR-2428545, ECCS-2344658, and ECCS-2139692).

10.1126/science.aec0463

Department of Physics, Carnegie Mellon University, Pittsburgh, PA, USA.
Email: sufeis@andrew.cmu.edu

POLICY FORUM

INDUSTRIAL DECARBONIZATION

Mobilizing capital and technology for a clean aviation industry

Research managers and investors need better approaches to weigh risks and potential benefits of scalable, transformational technologies

David G. Victor[1,2,3], Thomas Conlon[4], Philipp Goedeking[5], Andreas W. Schäfer[6]

Slashing warming pollution to the near-zero levels needed to stop global climate change will require massive investments in nearly every industry. However, nearly all of today's clean industrial investment is focused on a few sectors where technologies and business strategies are relatively mature and thus risks are relatively low: renewable energy, electric grids with battery storage, and electric vehicles (*1*). In other industries that lag behind, investment in new technologies and business practices involves much larger risks that are harder for investors to fathom. The aviation sector is emblematic of the challenge. A new approach is needed that is more realistic about what is achievable and better aligned with how real-world investors balance risk and reward. We outline an approach that could help guide a pioneer group of research and development (R&D) programs alongside investors and airlines who are motivated to deploy new technology.

Today about 3% of global emissions come from flying, a fraction that is rising as a wealthier and bigger global population demands more air travel. Jet airplanes mainly affect the climate from carbon dioxide (CO_2) pollution when they burn fuel and also by forming contrails from the water and soot that aircraft engines inject into the atmosphere at altitude. Like many other industrial sectors that are often called "hard to abate," such as steel, cement, and plastics production, aviation faces a multitude of technological and business challenges. Moreover, the industry's main product—passengers moved at a distance—is a fiercely competitive commodity. No airline or equipment manufacturer dares break much from the status quo lest it suffer economic losses. Yet, some are still motivated to take the lead and cut their impact on the climate.

Some of what is needed to create a clean technology in aviation is familiar: Governments, often working with industry, must support R&D to increase the supply of commercially viable new technologies such as new kinds of fuels and propulsion systems (*2*). Even with more R&D, however, a cleaner aviation industry will not emerge unless investors also have stronger incentives to back promising new technologies and apply them at global scale. About $1 trillion will be invested in the sector over the next decade—mainly by airlines and aircraft lessors as they acquire new airframes and engines (i.e., aircraft). New gear will make the global fleet a bit more efficient and thus slow the growth in emissions. A few airlines are also purchasing modest quantities of what's been cleverly branded as sustainable aviation fuel (SAF). Most SAF today, which comes from recycled fats and oils from the food industry, has lower–life cycle CO_2 emissions compared with conventional jet fuel. However, today's pathways for making SAF do not scale to the large volumes required to affect the whole aviation industry; to what extent new kinds of SAF from different feedstocks and production methods scale is also unknowable today.

Truly transforming the aviation sector to slash its impact on the climate will require much more than just tinkering at the margins with the minimal changes in aircraft and fuels. Bigger changes imply bigger risks, something that the purveyors of capital are wary about because they, too, compete in a market that abhors taking on too much risk with no extra reward. The result is that nearly all investors who are keen to advance sustainability operate mainly by screening out whole classes of investment (notably anything proximate to fossil fuels) rather than seeking investments that could be of high value to the climate. Green-minded investors have also largely avoided aviation because slow progress in inventing new fuels and technologies all but guarantees that the sector will not be on track to slash emissions to net zero by 2050, a widely discussed goal.

DECOUPLING TRAFFIC AND EMISSIONS

As global attention to climate change approaches its fourth decade, how could investors not have a workable framework for use in a sector as important as aviation? Some of the answer lies with the lack of clear incentives to cut emissions. In addition, policy debates around climate change, typically focused on the long term and on the geophysical goal of net zero emissions, have paid less attention to the practicality of how real investments alter industrial practices and emissions. When ambitious goals are set far into the future—such as net zero emissions by 2050—it is easy to gloss over all the contingencies and uncertainties that will determine exactly whether and how today's industry will transform into a substantially greener future. Through roadmapping exercises that focus on what might be needed to meet bold distant goals, it is easy (and dangerous) to confuse hypothetical needs with the practical realities of which technologies and projects will attract capital at large scale.

In some sectors, like semiconductors, roadmapping has played a profound role in guiding policy and industrial investment because key technological challenges are well known. The questions asked of roadmappers are tractable, such as where and how to invest in new production facilities. By contrast, nearly all the key questions in aviation—such as the cost and volume of potential SAF production beyond today's simple (but unscalable) methods, or the efficacy of schemes to limit aircraft contrails—are essentially unknowable without field experiments at scale (*2*).

In aviation one of the leading roadmappers has been the Science-Based Targets Initiative (SBTI) (*3*), which sets roadmaps based on scenarios provided by the International Energy Agency (IEA) (*4*), and International Council on Clean Transportation (ICCT) (*5*). Others include the Mission Possible Partnership (MPP) (*6*), Air Transport Action Group's Waypoint 2050 (*7*), the International Air Transport Association (*8*), and an academic team (UMI) (*9*) . But although it may be relatively easy for such exercises to imagine actions that could unlock green futures, it's much harder to see where and how investors will back them. One intuition from such scenarios, however, can help frame how to think about the future. A look across several leading aviation scenarios reveals a similar pattern (see the first figure). As demand for air travel rises, CO_2 emissions increase, at first, and then must plummet to meet ambitious distant goals. We use the base year of 2019—the last full year before the global pandemic wreaked havoc on the industry that is just now recovering. One advantage of this approach is that it allows direct measurement of the "decoupling" of what ails the industry

(emissions) from what yields profit and investment (passenger travel). In principle, decoupling should be a straightforward concept for the industry to adopt because it distinguishes what firms are optimized to do (grow market share and profitability) from the additional goal of cutting emissions. A focus on decoupling also makes it easy to measure progress (more decoupling) without indexing progress to absolute emission goals that can't yet be met (net zero emissions).

In many other sectors and in many countries, such decoupling charts have already been used, yet even though every major airline is already collecting data that make it possible to calculate its version of the first figure, it's far from clear whether or how any might be using such data for that purpose (*10*). This rhyming pattern of rising and then steeply falling emissions helps explain why roadmapping is deceptively attractive to industry leaders, and their investors, inclined to greenery. Nearly all investment in aviation today is focused on new aircraft whose financial lifetimes will be dominated by the next two decades or so—the rising part of the emissions curve on the first figure. Replacing today's fleet with more efficient new gear doesn't add much risk (except when new aircraft or engines don't work as expected or are delivered many years late) to an industry whose largest costs include fuel. Moreover, a few airlines are spending modest additional funds to purchase SAF, which helps them bring emission intensity down a bit more. Even with most of today's SAF coming from easily identified sources, most competitive airlines won't bear the cost. Few airlines are signing up to buy even bigger volumes of new kinds of SAF that are made from more scalable yet even costlier feedstocks and production methods, such as cellulosic biomass or synthesis with clean hydrogen.

The result is that no major airline is decoupling fast enough today—modest improvements in aircraft fuel efficiency are being more than offset by growth in airline services—and no major airline or major investor is holding itself accountable in terms of decoupling. Hypothetical roadmaps run to 2050 even though nobody in the industry appears to have much of a plan for equipment or SAF beyond the middle 2030s. Even regulations that, in theory, will require more SAF aren't reliable guides for what will actually happen in the future. For example, ReFuelEU mandates in the European Union (EU; the region that is most committed to adopting cleaner fuels) suggest SAF blending ratios through 2050, but a review in 2027 and every 4 years thereafter will determine whether quotas can be kept in place. If the rules become inconvenient, policy-makers will adjust them.

Roadmaps to net zero were supposed to guide the industry to clean futures, but a crisis is now brewing as the industry's leaders learn that what will happen in the real world of investable projects is different from what is needed to meet distant goals (*11*). The process of abandoning impractical goals began last year when Air New Zealand, which previously had one of the boldest emission goals in the industry, scaled back its aspirations for 2050 and even 2030 because they were infeasible to honor. (Air New Zealand now has a soft commitment to transparency but is avoiding specific targets.) Most other leaders in industry know they will be forced to rethink as well (*11*). Of the 33 airlines in the SBTi scheme, nearly a quarter have removed their short-term (before 2035) targets and just one airline has remained committed to a long-term target.

TARGETING R&D

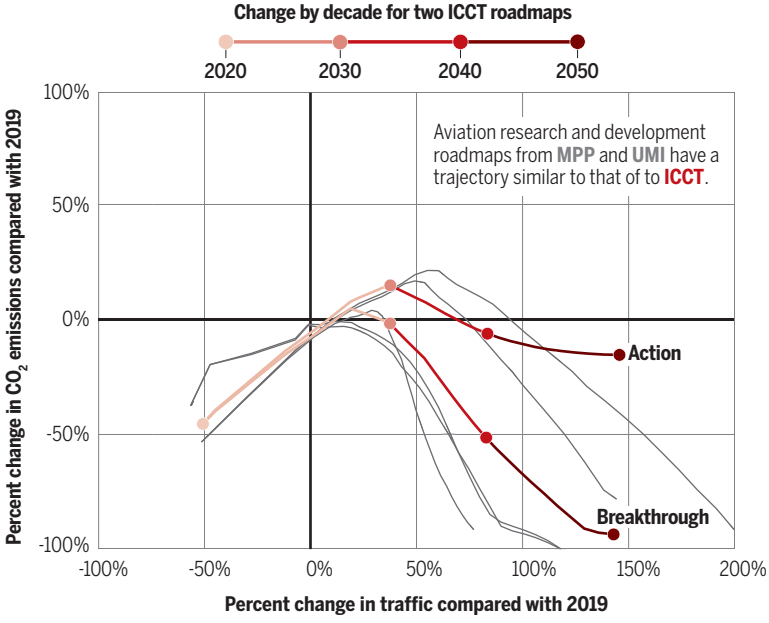
Creating a much cleaner aviation industry requires a much more active focus on root causes: the inadequate supply of radically cleaner technological systems, and the weak incentives for investors to adopt those technologies. Central to both these problems is risk. In a capital-intensive industry marked by cut-throat commodity competition, nobody will embrace those risks until there are new methods for supporting and rewarding them.

A key first step has long been central to the climate policy debate: boosting the supply of new ideas. A few governments are aiming to expand pipelines of viable new ideas and technologies. More will be needed, and the total level of investment must be much larger—but the key problem is that R&D managers, like most others in aviation, are prone to overweight familiar and less risky concepts (*2*). Yet, the goal of achieving near-zero carbon intensity across aviation suggests that the sector needs to look at radically different ideas alongside familiar ones. For example, some ideas for zero-carbon fuels can be adopted without much modification for use in today's aircraft, such as synthetic fuels from cellulosic biomass. Others, such as hydrogen, would require substantial changes to equipment and fueling infrastructure. To make the most of cleaner fuels will also require highly fuel-efficient aircraft designs such as flying wings that, for commercial aviation, remain risky and unfamiliar. The list of R&D needs grows beyond that to include much better understanding of contrails and contrail avoidance. It may also require electric aircraft platforms and new systems for certifying their safety.

Today's R&D programs don't adequately reflect this needed diversity. For example, the EU CleanSky research programs, starting in 2008 with an annual budget of €1.6 billion, and increasing to €4 billion in subsequent periods, have focused nearly exclusively on technologies that can be integrated easily in next-generation aircraft. The successor to CleanSky, the Clean Aviation program, gave some attention to hydrogen aircraft in 2022 but with the potentially unworkable aspiration of supporting hydrogen aircraft that might enter into service in the mid-2030s. Other parts of the EU R&D program support more radical breakthrough technologies—for example, the European Innovation Council with an annual budget of €1.4 billion, but no substantial-scale aviation decarbonization technology projects have been funded. The same is true of other national research

A familiar pattern across aviation roadmaps

Under seven leading roadmaps, as demand for air travel demand grows, so do carbon dioxide (CO_2) emissions initially, before dropping to meet longer-term ambitions. Emissions and traffic data use 2019 as a baseline. The roadmaps shown were developed by the International Council on Clean Transportation (ICCT), Mission Possible Partnership (MPP), and a team from University College London, the Massachusetts Institute of Technology, and Imperial College London (UMI). Traffic is measured as the number of metric tonnes of freight and passengers, times the distance flown, per year ("revenue tonne-kilometers"). See supplementary materials.



CREDITS: (GRAPHIC) V. PENNEY/*SCIENCE*; (DATA) INTERNATIONAL COUNCIL ON CLEAN TRANSPORTATION; MISSION POSSIBLE PARTNERSHIP; L. DRAY *ET AL.*, *NAT. CLIM. CHANGE* **12**, 956 (2022)

programs, such as in the United States, where government roadmaps put particularly heavy emphasis on familiar topics such as SAF. In practice, most SAF investment has gone into fuels recycled from waste oils and fats—a readily available feedstock that won't scale much and thus can't help the global industry.

A recrafted version of the first figure can help research managers assign quantitative weight to new technologies and systems that allow for the greatest reduction of emissions if applied at scale. We show two typical trajectories from future emissions scenarios—one highly optimistic and the other more modest (see the second figure). In our stylized example, we also mark grid cells with scores—what we call an "aviation sustainability index" (ASI)—that are proportional to the deviation from the diagonal. Analogous to credit ratings, the scores are set in bands, an approach that helps stabilize volatility in index ratings. An aviation sustainability finance group is already applying such an approach in practice and refining it with an eye to balancing granularity with complexity (*12*).

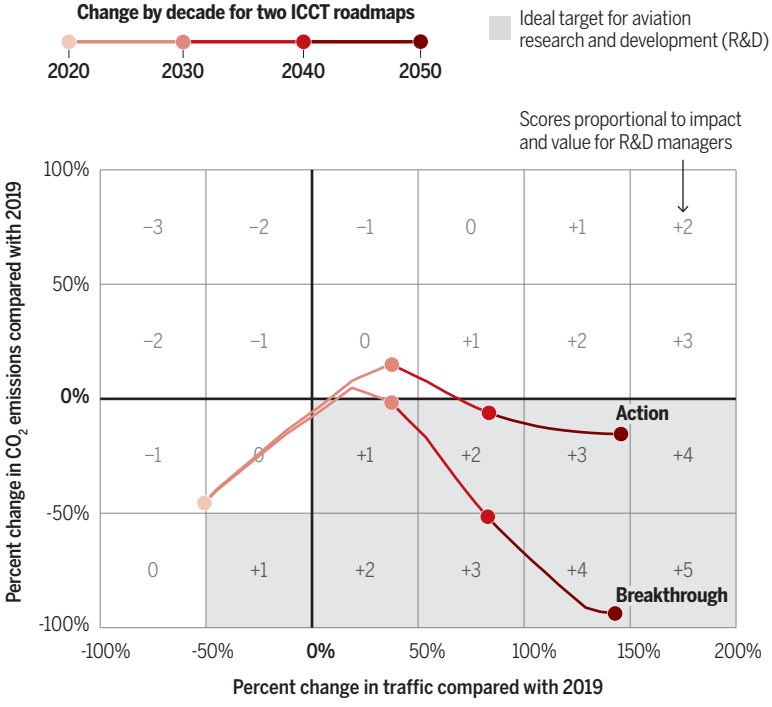
A portfolio of R&D investments should include an array of technologies that, if scaled, would lead to different-scored outcomes. Doing this will require that R&D managers also do something that any good program already does: Assess the probability that each innovation they back will scale industry-wide, with associated benefits for industry-wide impacts on the climate. Along the way, R&D managers will need to update their estimates and uncertainties as they learn which R&D bets are promising. And they will need to compare investments that affect emissions (notably CO_2) with other climate-beneficial actions such as learning how to avoid contrails. Managing R&D is a process beset by assumptions and counterfactuals—and the need for active management so that bets can be adjusted individually and as a portfolio (*2*).

A strategic approach to R&D informed by the second figure will, for the most part, lead managers to downweight lower value, familiar technologies (e.g., conventional SAF). The beneficiaries in R&D portfolios could be riskier investments, such as in new forms of SAF, programs to slash contrails, or innovative aircraft designs such as those that use electricity for portions of the flight—all areas where modest R&D programs exist today but are small when compared with more familiar and less risky innovations.

Of course, the world must also invest more in early-stage R&D. A few places have large R&D programs—predominately the UK, EU, and US, although the latter's may be bruised or eviscerated by the Trump administration—but the total effort lags. Philanthropy, which is important in some fields, is practically nonexistent in clean aviation. Along with more investment, the planet and industry would also benefit from more international cooperation because the market for clean aviation technology is global and no country or philanthropist can supply sufficient investment on its own. This cooperation must be led by the parties most committed to investment. Those first-mover countries, firms, and philanthropists might task the Clean Energy Ministerial to help convene the key players and perform an assessment of the overall global effort using the approach outlined in the second figure. Mission Innovation, another group with similar aspirations and overlapping membership, also has related activities under way. There is not a shortage of institutions that could help improve cooperation; what's been missing is a strategy that starts with a small club of the most highly motivated supporters of this research and keen to make sure their investments generate most value (*2*).

A conceptual approach to scoring investment opportunities

Innovations and investments leading to bigger departures from the diagonal of "0" scores in the grid reflect greater potential for emissions reductions. For example, changes consistent with the International Council on Clean Transportation (ICCT) "Breakthrough" scenario nearly completely decarbonize the aviation sector in 2050. Actions aligned with the less optimistic "Action" scenario reflect continuation of refleeting and other modest technological changes consistent with current patterns. The figure aligns with an approach developed by (*12*). Traffic is measured as the number of metric tonnes of freight and passengers, times the distance flown, per year ("revenue tonne-kilometers"). See supplementary materials.



GUIDING CAPITAL TO CLEAN

Transforming the aviation industry will require more than just a bigger and more diverse supply of clean technologies. Demand is also essential. And like all industries that are capital-intensive, the adoption of new technologies hinges on private investment.

Vast amounts of capital are already flowing into aviation. Little of that capital, however, is earmarked for clean aviation. Nearly all investment in the sector, instead, backs projects that already align with standard commercial metrics of risk-adjusted return without regard for decoupling emissions. In 2024, for example, $80 billion was invested in new aircraft globally, but only a tiny proportion of that investment was contracted on sustainability-linked loan facilities (*13*). In an industry where capital allocations are fungible across nearly the whole global economy and the users of capital (airlines, mainly) are locked in cutthroat commodity competition, nobody wants to take much extra risk. That's now changing in Europe, mainly, as governments tighten rules on emissions from all industrial sources along with mandates to use cleaner fuels in aviation—policies that apply, increasingly, to foreign operators as well as Europe's own airlines. Forward-looking airlines are beginning to take more risk, but they remain cautious lest they undermine their competitiveness.

In the absence of powerful global incentives to adopt clean aviation technologies, it is still possible to unlock a lot of investment by working with the large and growing pools of capital and investors who are willing to take some extra risk in exchange for environmental benefits. The European Investment Bank reports that in 2024 it supported 209 billion euros of sustainability-linked financing in all sectors, a number that is rising steeply every year despite the broader chaos in global financial markets.

CREDITS: (GRAPHIC) V. PENNEY/*SCIENCE*; (DATA) INTERNATIONAL COUNCIL ON CLEAN TRANSPORTATION

The challenge in aviation is the lack of any agreed framework for assessing investments and convincing funders that what they are doing is actually good for the planet. Much of the sustainability-linked finance has looked to roadmapping institutions like SBTi for guidance. As roadmapping goals and methods have drifted from the realities of what is achievable and how financial decisions get made, that approach has suffered. Thus, today investors who want to be sure they are green are adopting a different, perverse strategy: screening out any investments linked to fossil fuels. This feel-good approach to greenery excludes the places, like aviation, where more investment could yield huge benefits.

Doing better requires a framework that allows sustainability-minded investors to show how their extra tolerance for cost and risk enables a portfolio of investments that actually help the industry become greener than the status quo. Here the concept from the second figure can be helpful because the ASI value is an estimate of that deviation. Using data that all major airlines already collect, along with estimates of the future impacts of each investment on climate impacts, a target ASI can be estimated for single investments (e.g., a fleet of new efficient aircraft for an airline, a new SAF supply contract) or suites of projects aimed at spreading innovative technologies and practices across the whole industry. When seeking green financing, an airline or other project developer would compute a target ASI score, a first step to unlocking access to greener capital. As the project advances and meets performance milestones, it would be re-rated, opening up financial rewards, akin to other traded debt instruments.

As more investments use this approach, the market will help sort out the exact rewards for different kinds of risks and uncertainties associated with ASI targets—along with the rewards for successfully meeting those targets. Today's market, however, can't sort out those values because there is no agreed way to claim a benefit from a sustainability investment. Using this approach, we expect, will allow the greenest investors to set minimum ASI thresholds, akin to what is already done widely when securities and debt are rated—ratings above a certain threshold are "investment grade," which lowers the cost of borrowing typically about a whole percentage point and unlocks capital from investors who are required to back only bona fide high-quality investments. Because typical lending costs today are about 5 to 6.5%, a percentage point reduction can have a huge impact on investment patterns in this capital-intensive sector.

This approach can help sift "greenwashing" from transformative investments. Today's tools for sustainability lending link the terms of a loan (e.g., interest rates) to the achievement of sustainability targets, and include penalties for failure to perform and almost no positive incentives to do better—with the result that investors set goals that lack ambition and thus are easy to honor (*14*). Green investors offer few real financial advantages—perhaps a few hundredths of a percentage point in interest rates—and the firms that deploy that capital offer, in exchange, few real environmental advantages. Many of today's investments make aviation only a bit more efficient even as traffic grows—projects that would get nearly "0" ASI scores and thus earn no extra green reward, which is consistent with how many bankers view them. Big departures from the diagonal in the second figure yield the biggest absolute reductions in emissions and should earn the greatest extra value.

There is tremendous and growing pressure on lenders (e.g., banks), investment managers, and borrowers (e.g., airlines) to show that they are moving beyond the status quo—especially in places, such as Europe, where action on climate change remains robust. By making ASI scores transparent, and by rescoring as projects are implemented in the real world, it will be easier for lenders to guide their clean capital to investments that yield clean value. And by presenting a whole portfolio of ASI-guided investments, any investor or airline can also more easily explain to its constituents how its overall effort leads to decoupling over time.

In practice, ASI's can have great value when used with funding mechanisms that blend private capital with public funds—the former is available at massive scale, whereas the latter helps absorb extra risks. Early applications could include the risk-sharing approach to investing in green innovation led by the European Investment Bank (EIB) and other publicly minded investment banks that blend public funds with private investment. As with the index we propose, the key to the EIB's approach is identifying places where high-value investments carry higher risks that, through a blend of private and public finance, need extra support. A clean aviation investment strategy, perhaps also organized by EIB and focused on the European market to get started, could have similar effects focused on aviation.

The solutions needed in aviation are, in part, specific to the sector. Barriers to entry are high because regulatory certification procedures are highly risk averse—a problem magnified by the duopoly in major airframe manufactures. But in many respects, the challenges in this sector are shared across other "hard to abate" sectors such as shipping, heavy freight, steel, and others. All face the common challenge that there are many ideas for new technologies and business models, but none gets sufficient traction because investors don't know which vision is most credible; there are few rewards for taking bigger risks. Improving the supply of new technologies in tandem with ability to finance demand will unlock a cleaner future. □

REFERENCES AND NOTES

1. N. E. F. Bloomberg, "Energy transition investment trends 2025" (2025); https://about.bnef.com/energy-transition-investment/.
2. C. F. Sabel, D. G. Victor, *Fixing the Climate. Strategies for an Uncertain World* (Princeton Univ. Press, 2022).
3. SBTi, "Technical Report: The SBTi Interim 1. 5°C Sector Pathway for Aviation" (2023); https://sciencebasedtargets.org/resources/files/1.5C-Aviation-Interim-Technical-Report-Final.pdf.
4. IEA, "Net Zero Roadmap: A Global Pathway to Keep the 1.5 °C Goal in Reach" (2023); https://www.iea.org/reports/aviation-and-shipping.
5. ICCT, "Vision 2050. Aligning aviation with the Paris Agreement" (2022); https://theicct.org/wp-content/uploads/2022/06/Aviation-2050_report_final_v2.pdf.
6. MPP, "Making net-zero aviation possible. An industry-backed, 1.5°C-aligned transition strategy" (2022); https://3stepsolutions.s3-accelerate.amazonaws.com/assets/custom/010856/downloads/Making-Net-Zero-Aviation-possible.pdf.
7. Air Transport Action Group, "Balancing growth in connectivity with a comprehensive global air transport response to the climate emergency: A vision of net-zero aviation by mid-century" (2021); https://aviationbenefits.org/media/167417/w2050_v2021_27sept_full.pdf.
8. IATA, "Finance. Net Zero CO2 Emissions Roadmap (2024); https://www.iata.org/contentassets/8d19e716636a47c184e7221c77563c93/finance-net-zero-roadmap.pdf.
9. L. Dray *et al*., *Nat. Clim. Change* **12**, 956 (2022).
10. P. Tapio, *Transp. Policy* **12**, 137 (2005).
11. D. G. Victor, J. Dutra, "How to "blink responsibly" on unrealistic climate goals - and set a course for action" (2024); https://www.weforum.org/agenda/2024/10/climate-emissions-goals-blinking/.
12. impact e.V, "The milestones to decarbonize aviation" (2023); https://impact-on-sustainable-aviation.org/shared-files/1023/?Impact-White-Paper_Milestone-Concept_February23-2.pdf.
13. Boeing, "Commercial Aircraft Finance Market Outlook 2025" (2025); https://www.boeing.com/commercial/market/current-aircraft-financing-market#overview.
14. P. Feldhütter, K. Halskov, A. Krebbers, *J. Financ. Econ.* **162**, 103944 (2024).

ACKNOWLEDGMENTS

For extensive discussions about the interaction between aviation and finance we thank U. Ziegler, H. Kanters, J. Schirrmacher, M. Halaby, B. Monteleone, R. Miller, and C. Hügel. We are also grateful to participants at the IMPACT Forum in London, December 2024, where we presented and discussed these ideas. A huge thanks, too, to several reviewers for their insistence on clarity.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adu2458

10.1126/science.adu2458

[1] School of Global Policy and Strategy, University of California San Diego, La Jolla, CA, USA. [2]Division of Climate, Atmospheric Science and Physical Oceanography, Scripps Institution of Oceanography, La Jolla, CA, USA. [3]Foreign Policy Program, The Brookings Institution, Washington, DC, USA . [4]Graduate School of Business, University College Dublin, Dublin, Ireland. [5]Faculty of Law, Management and Economics, Johannes Gutenberg University, Mainz, Germany. [6]The Bartlett School of Environment, Energy and Resources, University College London, London, UK. Email: david.victor@ucsd.edu



BOOKS ET AL.

SOCIAL SCIENCE

Knowledge for two

A psychologist explores common knowledge and coordination

Jeremy Goodman and **Chaz Firestone**

Steven Pinker's new book, *When Everyone Knows That Everyone Knows…*, is about taking knowledge to infinity. Specifically, it is about common knowledge—not in the colloquial sense of facts we all take for granted but in a technical sense, as studied by philosophers, economists, linguists, game theorists, and political scientists since the 1960s. Common knowledge in this sense is infinitely interwoven interpersonal knowledge. For Alice and Bob to have common knowledge that it is raining, for example, it is not enough for them both to know it; they must also know that the other knows it, know that the other knows they know it, and so on, literally ad infinitum.



When Everyone Knows That Everyone Knows…
Steven Pinker
Scribner, 2025. 384 pp.

Pinker's excellent book brings psychology into the fold, marshaling decades of research on how "recursive mentalizing" (Pinker's term for reasoning about others' reasoning) shapes everyday behavior. In it, he argues that we regularly achieve common knowledge in all its infinite glory and that it is "a key to understanding human social life"—from meeting for coffee to successful politicking. "Two people really do need to think an infinite vortex of 'I know that he knows that I know that he knows…' thoughts to be sure they will coordinate their plans," he writes. "Two or three, or any finite number, aren't enough."

Consider a predicament called a "stag hunt" in game theory. It is lunchtime for Alice and Bob, who could each eat at their respective homes or head to the café. They are happiest eating together but do not want to risk eating alone at the café. Will they meet up? Merely knowing that they are happiest eating together might not be enough: If Alice doesn't know that Bob knows she prefers eating together, she'll worry that he'll play it safe and stay home; this worry might make her stay home herself. Without common knowledge, such doubts threaten to stymie coordination. Because coordinated behavior is commonplace, Pinker concludes that common knowledge is too.

Pinker anticipates, and deftly rebuts, numerous objections to this thesis—for example, can infinite layers of knowledge really fit in finite skulls? But the book leaves unanswered a fundamental challenge to the very possibility of common knowledge (let alone its prevalence). Epistemology—the philosophical study of knowledge—has seen an explosion of work on higher-order knowledge, provoked by Timothy Williamson's agenda-setting treatment (*1*). Williamson contends that we never have infinitely iterated knowledge of anything, even by ourselves (knowing that you know that you know…). More recently, Harvey Lederman (*2*) has argued that, even if we do have infinitely iterated knowledge as individuals, infinitely iterated knowledge between people—that is, common knowledge—remains unachievable and so cannot be what underwrites successful coordination.

Suppose Alice and Bob are neighbors who have arranged to meet for coffee on Saturday as long as it is above freezing (at least 33°F). Their smartphones have different weather apps, but they know that both apps are accurate to within a degree. On Saturday morning, they each privately check their phones and see 34°; so, they both know it is above freezing. But because they know that their apps could be a degree off, for all Alice knows, the actual temperature is 33°. From her perspective then, Bob's phone might read 33° or even 32°, and he wouldn't know it is above freezing. So, Alice and Bob lack common knowledge that it is above freezing.

The kicker is that this reasoning works at any temperature. Suppose Alice's and Bob's apps both read 60°. It is obviously above freezing, and there is no question that coffee is on. Do they have common knowledge now? No. If Alice's app reads 60°, then for all she knows it's really 59° outside and Bob's app reads 58°; in which case, for all Bob knows it's really 57° and Alice's app reads 56°; and so on, down to freezing.

This case has two morals. First, common knowledge—in the infinitary sense Pinker insists upon—may be downright unachievable. Lederman's challenge is both powerful and general: Even if Alice and Bob stand together in 60° air, how cool they feel may vary. For all Alice knows, it feels a degree cooler to Bob than to her, in which case, for all Bob knows it feels cooler still to Alice, and so on. By phone or by feel, they will lack common knowledge that it is above freezing. Second, common knowledge is not required for seamless coordination: If Alice's and Bob's apps each read 60°, they will surely meet.

Pinker does not deny that such coordination is possible; indeed, the book includes a long discussion of how salience allows us to coordinate by making educated guesses about what others will do. But Alice and Bob's coordination is not mere guesswork: At 60°, each knows that the other will show up. This knowledge is not mysterious: It simply applies an observed pattern of behavior—that the other person keeps their plans—to the case at hand. It only seems mysterious if we demand an explanation in terms of recursive mentalizing. While the book lucidly illustrates how social dynamics are often shaped by reasoning about reasoning, such demonstrations fall short of establishing its central thesis: that common knowledge is the key to coordination. □

REFERENCES AND NOTES

1. T. Williamson, *Knowledge and Its Limits* (Oxford Univ. Press, 2000).
2. H. Lederman, *Mind* **127**, 1069 (2018).

10.1126/science.aeb1081

The reviewers are at the Department of Philosophy and the Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD, USA. Email: jgoodm32@jh.edu; chaz@jhu.edu

ILLUSTRATION: MEROVINGIAN/ISTOCK PHOTO

CLASSICS REVISITED

A radically organic worldview turns 100

A philosopher's seminal book traced how science shaped the mentality of the modern world

Àlex Gómez-Marín

Alfred North Whitehead
The MacMillan Company, 1925. 212 pp.

SCIENCE AND THE MODERN WORLD | No modern thoughtscape makes sense except in light of the mental atmosphere brought about by science over the past 400 years. The English mathematician and philosopher Alfred North Whitehead (1861–1947) traced the origins and developments of such a "climate of thought" in *Science and the Modern World*, a great series of lectures published in 1925, whose centennial is worth celebrating. In an extraordinarily comprehensive effort, Whitehead explains the key conceptual movements of science from Galileo to Einstein. Century after century, he masterfully highlights a living history of world-shaping ideas, revealing the deep habits and unfolding trends of the scientific mentality.

Science was born in the 17th century as a reaction to medieval thought that nevertheless embraced its belief in a natural order. Its forerunners admirably blended rationality and empiricism, balancing powerful abstractions with attention to concrete facts. Mathematical refinements and mechanical insights propelled science's success throughout the 18th and 19th centuries, led by pioneering French mathematicians such as Pierre-Simon Laplace and Joseph-Louis Lagrange and Italian physiologists including Luigi Galvani and Alessandro Volta.



English mathematician and philosopher Alfred North Whitehead traced key conceptual movements of science in his influential work.

The 19th century was marked by what Whitehead calls "the invention of the method of invention," together with the professionalization of science. These advances were met with the "Romantic reaction," opposing the mechanical view of life and affirming the intrinsic reality of value. Whitehead underscores key scientific ideas in this period: the notion of fields leading to the electromagnetic synthesis of James Clerk Maxwell; the concept of atomicity successfully deployed in both chemistry and biology through molecules and cells, respectively; and the gestation of the theory of evolution. He then turns to the 20th-century revolutions in modern physics and their counterintuitive nature, which his "organic philosophy" seeks to accommodate. His take on quantum physics and relativity is original and profound but remains obscure and untrodden.

What is so remarkable about science? For Whitehead, it is empirical observation coupled with mathematical reasoning. More than its "method," he argues, it is the scientific "mentality" that has profoundly influenced our lives and morphed societies for generations. He compares the acts of "great conquerors" with the ideas of thinkers, "individually powerless, but ultimately the rulers of the world."

Whitehead celebrates science's great accomplishments but critiques its abstractions when they go too far, especially when they are mistaken as concrete. The most pervasive and pernicious scientific abstraction of all, he argues, is "simple location": the idea that there is merely motion of material in a mechanical universe, namely, a world made of independent bits of matter distinctly located in space and time. The notion of mass is actually the great pearl in the crown of scientific materialism but also its Achilles' heel, as it afforded science with the basis for measurable and regular laws at the price of adopting an "unbelievable" idea, writes Whitehead, one that does violence against the very facts of our immediate experience and also leads to the mechanization and, ultimately, the destruction of the natural world.

In *Science and the Modern World*, Whitehead also offers hints of his own radically organic view of nature, anticipating the grand philosophical scheme that would be published 4 years later in his magnum opus, *Process and Reality*. He proposes to make relations no less foundational than the things related, replacing "substances" with "events" as the real constituents of the world. His cosmology naturally accommodated (and even anticipated) in real time the radical transformations in humanity's conceptions of space, time, matter, and measurement that occurred at the dawn of the 20th century.

Whitehead's thought is a century old, and yet it still seems to stream in from the future. His ideas are finally being caught up to in cutting-edge biology, which is currently enjoying an organicist renaissance: the crumbling of gene centrism, the recognition of organismal agency, a grasping of the processual nature of development, and a more sophisticated understanding of evolution. A similar subtle revolution is also going on in the cognitive sciences, where mind and life interpenetrate, and in consciousness studies, where the gulf between mind and matter is also being bridged and dissolved. Back to theoretical physics, Whitehead offers a path forward for a field stalled and captured by "theories of everything" so disconnected from facts and concrete human experience that they seem to have abandoned the realm of science altogether.

Thinking along with Whitehead, we can escape our ideological "groove." Nothing seems more urgent in our polarized, postmodern, artificial intelligence–seized age. □

10.1126/science.adz5359

The reviewer is at the Instituto de Neurociencias, Consejo Superior de Investigaciones Científicas–Universidad Miguel Hernández de Elche, Alicante, Spain.
Email: agomezmarin@gmail.com

PHOTO: HERITAGE IMAGE PARTNERSHIP LTD/ALAMY



LETTERS

Edited by
Jennifer Sills

Green algae (*Ulva prolifera*) covers the waters of the Yellow Sea in China's Shandong Province.

PHOTO: VCG/GETTY IMAGES

Stemming green tides in the Yellow Sea

Green tides typically composed of the alga *Ulva prolifera* have surged in frequency and extent over the past four decades, affecting more than 37 countries across the Asia-Pacific, Europe, and the Americas (*1*). For nearly two decades, China's Yellow Sea has been battling the world's largest and most persistent green tides (*2*), which emerged again in June (*3*) and spread across the entire Shandong Peninsula (*4*). An estimated 1 million tonnes of green algae decompose annually in the Yellow Sea, releasing substantial amounts of nutrients and organic matter into the ecosystem (*5*). As dense *Ulva* mats decay, they consume large amounts of dissolved oxygen, disrupt benthic habitats, and release nutrients that reinforce a self-sustaining eutrophic cycle (*2*). China must work to prevent these recurring disturbances, which erode biodiversity, disrupt ecological services, foul beaches, and disrupt seasonal economies, imposing mounting costs on coastal tourism and management (*1*, *2*).

Green tides arise from the convergence of nutrient pollution, opportunistic algal growth, and favorable oceanographic conditions (*6*). The large-scale seaweed aquaculture along the Yellow Sea's southern coast makes the region particularly vulnerable. *U. prolifera* grows on seaweed rafts and then detaches each spring when the structures are dismantled for harvest (*2*, *7*). Carried northward by prevailing currents and seasonal winds, the floating mats encounter eutrophic, poorly flushed nearshore waters that fuel explosive growth.

To manage seasonal blooms, coastal authorities rely heavily on physical removal, a labor-intensive effort that provides only short-term relief. In 2023 alone, Qingdao authorities removed nearly 247,000 tonnes of *Ulva* from offshore waters, along with more than 6000 tonnes from beaches and shorelines (*8*). Efforts are underway to use *Ulva* as agricultural soil conditioners, nutritional supplements for livestock, and feedstocks for biogas production (*9*) and to enzymatically extract high-value ingredients from *Ulva* for use in food, cosmetics, and pharmaceuticals (*9*, *10*). However, such efforts are constrained by biomass variability, high logistical costs, and limited market infrastructure. Neither physical removal nor resource recovery addresses the underlying ecological toll of the blooms.

Instead, enduring solutions must begin with stringent controls on nitrogen and phosphorus inputs from agriculture, aquaculture, and wastewater, reinforced by stronger regulation and consistent oversight. Effective management also requires timely detection through integrated monitoring systems that combine satellite sensing with predictive modeling. To prevent these blooms from escalating into region-wide crises amid accelerating climate change, a systemic approach is imperative. Policy-makers, researchers, and industry stakeholders must collaborate to prevent green tides and safeguard the long-term resilience of coastal ecosystems.

Luhua Yang[1], Ruiyong Zhang[1], Li Zhuo[2], Bin-Bin Xie[2], Jianqun Lin[2], Shun Li[2]

[1]State Key Laboratory of Advanced Marine Materials, Laboratory of Marine Environmental Corrosion and Bio-fouling, Yellow Sea Ocean Observation and Research Station, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China. [2]State Key Laboratory of Microbial Technology, Shandong University, Qingdao, China. Email: shunli@sdu.edu.cn

REFERENCES AND NOTES

1. N. Ye *et al.*, *Ecol. Res.* **26**, 477 (2011).
2. Y. Zhang *et al.*, *Natl. Sci. Rev.* **6**, 825 (2019).
3. J. Song, J. Cao, L. Zhu, "Sea surface turns into a 'grassland'? Green tide outbreak hits parts of the coastal waters off Yantai, Shandong, experts explain the causes" [in Chinese], *CCTV News*, 4 July 2025; https://content-static.cctvnews.cctv.com/snow-book/index.html?item_id=8287347645440819835.
4. J. Fan, "Sea surface turns into grassland? Coastal waters in multiple regions of Shandong have turned into 'spinach soup'" [in Chinese], *Shangguan News (Wen Hui Daily)*, 10 July 2025; https://www.jfdaily.com/staticsg/res/html/web/newsDetail.html?id=944269&v=2.0&sid=67.
5. G. Zhang *et al.*, *Remote Sens.* **16**, 4794 (2024).
6. V. Smetacek *et al.*, *Nature* **504**, 84 (2013).
7. Z. Wang *et al.*, *Limnol. Oceanogr.* **60**, 1105 (2015).
8. J. He, "Green tide turns the sea surface into a 'grassland'? How can we combat *Ulva* more effectively?" [in Chinese], *CNR News Online*, 10 July 2023; https://china.cnr.cn/gdgg/20230710/t20230710_526324329.shtml.
9. L. C. Hofmann *et al.*, *Crit. Rev. Food Sci. Nutr.* **65**, 3728 (2025).
10. M. Fournière *et al.*, *Mar. Drugs* **19**, 156 (2021).

10.1126/science.aea7703

Australian algal blooms require national action

Harmful algal blooms are devastating marine life at an unprecedented scale across South Australia's Great Southern Reef (*1*) and

are expected to increase with climate change (*2*–*4*). A prolonged marine heat wave and calm seas have enabled toxic dinoflagellate algae (predominantly *Karenia mikimotoi*) to flourish in stratified coastal waters (*2*). Because harmful algal blooms affect society and environments across jurisdictional boundaries, Australia should coordinate a national management approach (*4*, *5*).

Akin to the prolonged disaster of a drought, but with ecological impacts similar to a bushfire (*6*), the harmful algal blooms have produced mass mortalities of diverse marine vertebrates and invertebrates (*7*) across hundreds of coastal kilometers, degraded the marine habitats that support them, and forced shellfish fisheries to close (*8*). However, unlike droughts, Australia has no agreement in place that ensures intergovernmental cooperation (*9*). Unlike bushfires, Australia's National Emergency Management Agency does not recognize harmful algae blooms as natural disasters. When Australia's "black summer" fires ravished eastern and southern coastal bushlands in 2019 and 2020, the Australian government responded with a nationally coordinated bushfire emergency response that included more than AUD 2.2 billion to support community and environmental recovery and improve future management (*10*), compared with just AUD 28 million in combined Federal and State Government funding for the *Karenia* algal bloom (*11*). Prolonged harmful algal blooms that cause immediate and long-term environmental and socioeconomic impacts require a national response similar to the ones in place for droughts and bushfires.

Australia's Great Southern Reef is a global biodiversity hotspot worth more than AUD 11.5 billion per year to the nation (*1*). More than 1000 citizen scientists have documented over 52,000 mortalities along the affected shoreline (*12*). This damage is likely just a fraction of the harm to the reef. Australia must prepare for future catastrophic harmful algal bloom events and marine heat waves. These events require a national response that streamlines efforts across local, state, and federal governments to rapidly assess the underwater impacts and identify future risks. Only a coordinated national strategy can improve long-term adaptation to and mitigation of Australia's marine disasters.

Dominic McAfee[1,2] and Brad Martin[3]

[1]School of Biological Sciences, The University of Adelaide, Adelaide, SA, Australia. [2]Environment Institute, The University of Adelaide, Adelaide, SA, Australia. [3]OzFish Unlimited, Ballina, NSW, Australia. Email: dominic.mcafee@adelaide.edu.au

REFERENCES AND NOTES

1. S. Bennett *et al*., *Mar. Freshw. Res*. **67**, 47 (2015).
2. S. A. Condie, E. C. Oliver, G. M. Hallegraeff, *Harmful Algae* **87**, 101628 (2019).
3. D. M. Anderson *et al*., *Harmful Algae* **102**, 101975 (2021).
4. S. K. Moore *et al*., *PLOS Clim.* **3**, e0000411 (2024).
5. M. D. Howard *et al*., *Integr. Environ. Assess*. **19**, 586 (2023).
6. P. Macdonald, "'Like an underwater bushfire': SA's marine algal bloom is still killing almost everything in its path," *The Conversation*, 12 June 2025.
7. J. W. Oh, S. S. C. Pushparaj, M. Muthu, J. Gopal, *Plants* **12**, 3936 (2023).
8. D. Keane, T. Kelsall, "Algal bloom 'disaster' prompts royal commission call as impact spreads to mussel farms," ABC News, 13 July 2025.
9. Australian Government, National Drought Agreement (2024); https://www.agriculture.gov.au/agriculture-land/farm-food-drought/drought/drought-policy/national-drought-agreement.
10. Australian Public Service Commission, "Black summer," in chap. 1 of "State of the service report 2019-20" (Australian Government, 2019); https://www.apsc.gov.au/state-service/state-service-report-2019-20/chapter-1-commitment-service/black-summer#_ftn3.
11. Government of South Australia, Algal bloom update, Grants and support (2025); https://www.algalbloom.sa.gov.au/support-from-government.
12. iNaturalist, SA marine mortality events 2025 (2025); www.inaturalist.org\\projects\\sa-marine-mortality-events-2025.

10.1126/science.aea3178

Chile's framework law erodes institutions

On 29 September, Chile passed Law No. 21,770, which eases the environmental permitting process for governmental and industrial projects. A response to the assumption that environmental permits slow down investment and stall the country's economic development, the law aims to shorten sectoral permitting times by 30 to 70% (*1*). However, the regulatory changes in the law undermine Chile's decades-long process of modernizing and strengthening its environmental institutions. To protect ecosystems that are critical to scientific research and climate action, Chile should implement strong secondary regulations and transparent oversight mechanisms.

Over the past three decades, Chile has transitioned from a framework of weak environmental institutions inherited from the civil-military dictatorship (*2*) to a more technical, robust, and internationally aligned governance model. This transformation included creating the Ministry of the Environment, improving the Environmental Impact Assessment System, and enacting the Framework Law on Climate Change (*3*–*5*). These advances enhanced the state's capacity to evaluate, prevent, and oversee the socioenvironmental impacts of investment projects.

Law No. 21,770 introduces a risk-based management approach that allows some permits to be replaced by sworn statements for activities deemed low risk, such as landfills, riverbed interventions, native forest management, and wastewater treatment. This policy weakens preventive oversight mechanisms and delegates key issues to secondary regulations. Such changes will erode legal certainty, expand administrative discretion, and defang regulatory tools built over decades.

The permitting changes threaten highly sensitive ecosystems, the protection of which is crucial for biodiversity conservation and climate mitigation. Chile is home to 88 of the world's 110 ecosystem types (*6*), making it essential to global biodiversity goals. Chilean ecosystems also play a vital role in climate regulation. In the Andes and Patagonia, *Sphagnum*-dominated peatlands and temperate forests act as effective carbon sinks, capturing twice as much carbon per hectare as the Amazon (*7*). Chile's ecological heritage includes the Valdivian rainforest, which is home to the world's oldest living tree, a *Fitzroya cupressoides* that is more than 5000 years old (*8*). Relaxing environmental permitting without rigorous assessment opens the door to cumulative and irreversible effects such as biodiversity loss, erosion, depletion of water reserves, and the weakening of carbon sinks.

Strong secondary regulations, standards applied by public agencies, and effective environmental oversight can mitigate the negative effects of the law. Regulatory bodies can reduce ecological risks by demanding thorough cumulative impact assessments and narrowing the definition of activities considered low risk. Scientific institutions can produce the knowledge and data needed to guide these decisions, and local communities and civil society groups can press for greater transparency, accountability, and protection of nature.

Iván Ojeda-Pereira[1,2]

[1]Department of Social Sciences, School of Engineering Sciences, Lappeenranta-Lahti University of Technology, Lappeenranta, Finland. [2]Institute of Sociology, Pontificia Universidad Católica de Chile, Santiago, Chile. Email: ilojeda@uc.cl

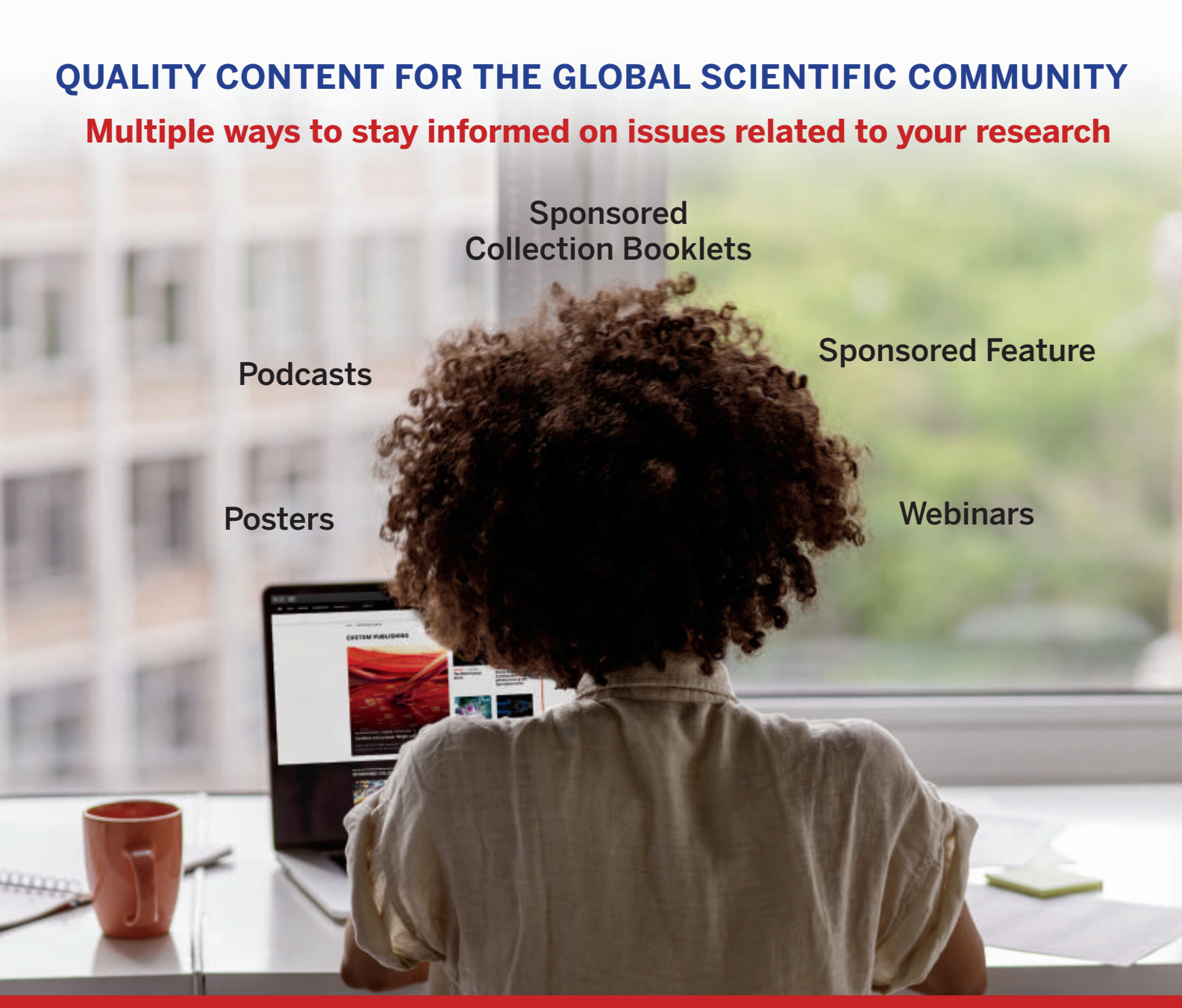
REFERENCES AND NOTES

1. Government of Chile, "Sectoral permit law approved: Timeframes for granting permits for investment projects to be reduced," Gob.cl (2025); https://www.gob.cl/en/news/sectoral-permit-law-approved-timeframes-for-granting-permits-for-investment-projects-to-be-reduced/.
2. D. Carruthers, *Third World Q.* **22**, 343 (2001).
3. A. Madariaga, *Bull. Lat. Am. Res.* **38**, 453 (2019).
4. A. Maillet, S. Carrasco, "Chile y el cambio climático. Pensar globalmente, actuar localmente" [in Spanish], C. Carrasco, Ed. (FES y Fundación La Alameda, 2019), pp. 66–92.
5. F. Campos-Medina, *Time Soc.* **28**, 1239 (2018).
6. D. A. Keith *et al*., *Nature* **610**, 513 (2022).
7. J. F. Perez-Quezada *et al*., *Austral Ecol.* **48**, 893 (2023).
8. G. Popkin, "Is the world's oldest tree growing in a ravine in Chile?" *Science*, 20 May 2022.

10.1126/science.aea4024

QUALITY CONTENT FOR THE GLOBAL SCIENTIFIC COMMUNITY

Multiple ways to stay informed on issues related to your research



Science
AAAS

Scan the code and start exploring the latest advances in science and technology innovation!

Science.org/custom-publishing

Brought to you by the *Science*/AAAS Custom Publishing Office.

Posters | **Podcasts** | **Sponsored Collection Booklets** | **Sponsored Feature** | **Webinars**

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**



PHOTO: MINORU MORIYAMA

ANIMAL SYMBIOSIS

Fungal obstructionism

Many invertebrates are preyed on by parasitoid wasps that lay their eggs in or on another organism's body for their hatchling larvae to feed upon. Nishino *et al.* discovered that females of a small dinidorid stinkbug have evolved a mechanism to help their offspring avoid this fate. The wasps have symbiotic organs on their hindlegs with a tympanum-like structure. The outer cuticle has pores through which the authors observed fungal hyphae emerging. Several fungal species are selected by the females, which notably include cordyceps that are often insect pathogens. When the female bugs lay their eggs, they rub the cultivated hyphae across the egg mass. The hyphae grow to envelope the eggs and physically exclude attentive parasitoid wasps until the bugs hatch. —Caroline Ash

Science p. 279, 10.1126/science.adp6699

Some stinkbug species, such as *Coridius chinensis* pictured here, use a fungal symbiosis to protect their eggs from parasitic wasps.

ORGANIC CHEMISTRY

Polypropionate prep

Polypropionates are a class of natural product derivatives that feature alternating hydroxyl and methyl stereocenters on a carbon backbone. Targeting particular stereochemical distributions is an enduring challenge in their synthetic preparation for pharmaceutical research. Nakao *et al*. report a versatile route exploiting asymmetric chromium-catalyzed allylation paired with photoredox activation of simple butene reagents and a boron-promoted ring opening of a cyclic precursor. The protocol can be applied consecutively with tunable stereochemical selectivity in each successive iteration. —Jake S. Yeston

Science p. 272, 10.1126/science.adz0686

INFECTIOUS DISEASE

If you build it, they won't come

Sociality is essential to many species, but it can come with the downside of increasing the transmission of disease. Many species have thus evolved strategies that limit disease transmission, even within highly social species. Leckie *et al*. tested whether the introduction of a pathogen through nest mate exposure altered patterns of nest construction in black garden ants. They found that the presence of individuals exposed to a pathogen led to architectural changes within the nest, such as an increase in the distance between entrances, and models showed that these changes reduced disease transmission. —Sacha Vignieri

Science p. 266, 10.1126/science.ads5930

ORGANIC CHEMISTRY

A porphyrin bracelet

Over the past decade, chemists have made steady progress in synthesizing a variety of carbon nanobelts. These band-shaped molecules consist of edge-sharing aromatic rings and resemble small cross-sectional fragments excised from carbon nanotubes. Rodríguez-Rubio *et al*. report a series of nanobelts composed of eight to 12 nickel-coordinating porphyrin macrocycles. They come together surprisingly simply through a coupling reaction promoted by additional nickel, followed by gold-induced oxidation. Spectroscopy revealed aromatic or antiaromatic electron behavior for belts with odd or even numbers of porphyrins, respectively. —Jake S. Yeston

Science p. 290, 10.1126/science.adz8900

PLANT ECOLOGY

Lower resilience to extreme droughts

Climate change is bringing more severe and longer-lasting droughts to many locations around the world. Some ecosystems have shown resilience to increasing drought, but this could change as droughts become more severe. Ohlert *et al*. assessed drought effects with a rainfall-exclusion experiment distributed across 74 grasslands and shrublands on six continents. On average, primary productivity declined by 29% in moderate-drought years, but effects did not compound over multiyear droughts. Extreme droughts (those falling outside the historic range of rainfall variability) caused much higher productivity declines and lower recovery, especially in drier locations and those with lower plant diversity. This study adds to growing evidence that future climate extremes will likely affect ecosystems in new and concerning ways. —Bianca Lopez

Science p. 284, 10.1126/science.ads8144

ANTHROPOLOGY

Human brain evolution and lead exposure

Severe exposure to environmental lead is a relatively recent phenomenon in human history that is known to cause permanent damage to the brain and nervous system, especially in children. In a study of several early hominin species ranging over a time span of 2 million years, Joannes-Boyau *et al.* showed that our ancestors were consistently exposed to lead, which could have affected the expression of the *FOXP2* gene that has been shown to be critical for the development of human speech and language acquisition. The authors postulate that positive natural selection has led to the replacement of the archaic variant of the *NOVA1* gene that disrupts *FOXP2* expression, thus offering a survival advantage to our hominin ancestors. —Mark Aldenderfer

Sci. Adv. (2025)
10.1126/sciadv.adr1524

ORGANIC CHEMISTRY

Spiroaspertrione A is synthesized

Spiroaspertrione A is a complex, polycyclic fungal natural product that has garnered interest in antibiotics research. Huang *et al*. now report a total synthesis of the compound in its natural stereoconfiguration starting from readily available precursors. Key steps included a Diels-Alder reaction to set the compound's characteristic spiro stereocenter, as well as a divinylcyclopropane rearrangement to establish an additional out-of-plane ring. The synthesis required 16 steps from a reported precursor and proceeded in an overall yield of 2.3%. —Jake S. Yeston

Science p. 261, 10.1126/science.adz7593

IMMUNE CHECKPOINTS

Phagocy-"no"-cis blockade

Inhibitory checkpoint blockades to increase macrophage antitumor responses have yielded disappointing therapeutic results. SIRPα is a key inhibitory receptor expressed by macrophages, and engagement of its ligand, CD47, which is expressed by tumor cells, prevents phagocytosis. However, antibody blockade of these SIRPα–CD47 trans interactions have achieved inconsistent clinical efficacy. Tang *et al*. revealed an additional mechanism whereby SIRPα suppressed phagocytosis and acted independently of CD47. Phagocytosis was prevented through cis interactions between SIRPα and CD18 expressed on the surface of macrophages. Dual blockade of SIRPα's cis (CD18) and trans (CD47) interactions potentiated phagocytic function and promoted antitumor immunity. —Leoma Bere

Sci. Immunol. (2025)
10.1126/scitranslmed.adv5085

IN OTHER JOURNALS

Edited by **Corinne Simonti** and **Jesse Smith**

CELL BIOLOGY

Breaking out and breaking in

Cerebral malaria is a serious complication of *Plasmodium falciparum* infection that compromises the blood-brain barrier (BBB). Piatti *et al.* bioengineered a microvascular model of the BBB using endothelial cells, pericytes, and astrocytes. Parasite products released during egress from infected red blood cells increased vascular permeability by down-regulating endothelial junctional and vascular development pathways. Furthermore, *P. falciparum* egress products activated inflammatory and antigen presentation pathways in all cells composing the BBB. Cytoadhesive *P. falciparum*–infected red blood cells led to locally confined disruptive effects. Ruxolitinib, an inhibitor of the inflammatory JAK-STAT pathway, prevented *P. falciparum*–induced microvascular leakage and may thus represent a therapeutic option to mitigate the complications of cerebral malaria. —Stella M. Hurtley

Nat. Commun. (2025)
10.1038/s41467-025-62514-2

IMMUNOBIOLOGY

KRAS and myeloid cells

KRAS is a gene frequently mutated in human cancers. Mutations in *KRAS* leave the gene stuck in the "on" position, where it sends continuous signals to cancer cells to grow and divide. The development of drugs that can switch KRAS "off" is an active area of biomedical research, and understanding the potential side effects is key. When Zamorano-Dominguez *et al.* genetically inactivated *KRAS* in adult mice, they found elevated numbers of myeloid immune cells. Systemic *KRAS* ablation led to the development of myelomonocytic metaplasia, a condition in which the bone marrow produces too many white blood cells and can lead to anemia, bleeding, and infections. These insights may shed light on the possible side effects associated with KRAS inhibitors. —Priscilla N. Kelly

Proc. Natl. Acad. Sci. U.S.A. (2025)
10.1073/pnas.2512404122

ENVIRONMENT

Is this river alive?

A pressing concern today is anthropogenic pollution of freshwaters. However, anthropogenic climate change can also create conditions that indirectly cause lake and river pollution. Sullivan *et al.* have investigated the source of the increasing turbidity and discoloration of the Salmon River watershed in Alaska that was first noticed in 2019. They found elevated levels of iron, sulfate, and toxic metals. The pollution now is so bad that once-clear streams have become unfavorable for life. The source of this pollution is climate change–associated permafrost thaw, which has exposed sulfide minerals in the underlying bedrock and created conditions resembling acid mine drainage. This phenomenon appears to be affecting many Arctic streams on similar sulfidic geology. —Caroline Ash

Proc. Natl. Acad. Sci. U.S.A. (2025)
10.1073/pnas.2425644122

ECONOMICS

Civil rights roots in military racism

Black men who faced racism while in the US Army during World War I were more likely to later become pioneers in the fight for civil rights. Using



DEEP-SEA BIOLOGY

Turning toxins into safe minerals

Deep-sea hydrothermal vents are among the harshest environments on Earth, with toxic levels of arsenic and hydrogen sulfide. Wang *et al.* studied the worm *Paralvinella hessleri*, which thrives right next to these vents, to understand how it survives such extremes. They discovered that the worm accumulates arsenic and sulfur inside its cells, where the two toxins combine to form tiny crystals of orpiment, a yellow mineral. By transforming poisons into a stable, harmless form, the worm can tolerate extreme conditions. This discovery highlights a unique survival mechanism and broadens our understanding of animal adaptation and biomineralization. —Di Jiang *PLOS Biol.* (2025) 10.1371/journal.pbio.3003291

The hydrothermal vent-dwelling worm *Paralvinella hessleri* survives toxic levels of arsenic and sulfur by converting them into yellow pigments.

millions of digitized records from the military and from the National Association for the Advancement of Colored People (NAACP), Ang and Chinoy exploited random selection through a draft lottery to show that Black draftees were more likely to later join the NAACP and be recognized as leaders in the Black community. Effects were pronounced among soldiers who were denied opportunities for training and promotion or who risked their lives while facing racist abuse, and were smaller among those who faced less discrimination while brigading with the French military. —Brad Wible

Q. J. Econ. (2025) 10.1093/qje/qjaf046

MANAGED RETREAT

Marching from the sea

Coastal communities are some of the most vulnerable regions to climate warming because of sea level rise and coastal storms, and many will be forced to move inland as a result. How have coastal settlements already responded to these threats? Xu *et al.* analyzed coastal migration trends by examining changes in nighttime lighting over the period from 1992 to 2019. They found that settlements have retreated inland in more than half of the regions due to increasing coastal hazards, insufficient protection by infrastructure, and less adaptive capacity. These observations should help us understand better when managed retreat will be required and the planning challenges involved. —Jesse Smith

Nat. Clim. Chang. (2025) 10.1038/s41558-025-02435-6

SYNTHETIC BIOLOGY

Engineering robust containment in the gut

There has been great interest in modulating or supplementing the gut microbiota with engineered microbial strains that might aid in treating disease or improving health. However, engineered microbes must be genetically contained so that they do not survive outside of the individual being treated and do not transfer genetic material to other microbes. Building on an earlier system in which a nuclease degrades plasmid DNA at temperatures below a certain threshold, Foo *et al.* engineered an additional layer of control that prevents nuclease inactivation by insertion mutations using an "entangled" gene architecture in which two genes overlap but use different reading frames. The system achieves stringent containment, with an escape frequency several orders of magnitude lower than previous systems. —Michael A. Funk

ACS Synth. Biol. (2025) 10.1021/acssynbio.5c00412

PHOTO: WANG H, ET AL., 2025, PLOS BIOLOGY, CC-BY 4.0

Edited by **Michael Funk**

PROTEOMICS

A cell-level look at proteins

Single-cell analysis is increasingly popular in biomedical research, but most single-cell studies are focused on transcriptomics. These messenger RNA (mRNA) analyses allow for the classification of cell types and changes in the cells over time, but they do not directly reflect cellular functions, which depend on proteins, and the abundance of a given mRNA often does not correlate with that of its corresponding protein. To help address this knowledge gap, Furtwängler *et al*. developed a mass spectrometry–based method for performing single-cell proteomics. They demonstrated their method on bone marrow samples from healthy human donors, providing biological insights into the process of differentiation of hematopoietic stem and progenitor cells. — Yevgeniya Nusinovich

Science p. 254, 10.1126/science.adr8785

LINEAGE TRACING

Tracing tumor evolution with PEtracer

All cells in an organism come from other cells. Accordingly, biologists have long sought to understand the dynamics of cell fate determination in development and disease. Koblan *et al*. developed PEtracer, a prime editing–based evolving lineage-tracing system that uses predefined edits to mark cell divisions over diverse spatial and temporal scales. PEtracer jointly profiles cell state and lineage in intact tissue using both single-cell sequencing and high-resolution, multimodal imaging. Using PEtracer in a syngeneic mouse model of cancer metastasis, the authors uncovered how the interplay between cell-intrinsic programs and spatial organization shapes cancer growth in three dimensions, leading to distinct programs of tumor evolution —Di Jiang

Science p. 255, 10.1126/science.adx3800

PIGMENTATION

One locus, many looks

Even when controlled by single genes, traits can still show great variability when multiple regulatory elements affect their expression. Lutgen *et al*. extensively characterized four species of wheatears, birds that are notable for their varied black or white mantles, throats, and necks. They found that although a single locus containing the known pigmentation gene *ASIP* segregates with these traits, each body part is affected by different regulatory variation in addition to a coding variant responsible for white feathers. The story is complicated further by the extensive hybridization of these species, leading to an old white haplotype spreading more broadly from one species, *Oenanthe melanoleuca*, and more recent mutations causing similar phenotypes in others. —Corinne Simonti

Science p. 256, 10.1126/science.ado8005

IMAGING

Bladeless sectioning for nanoscale imaging

High-resolution microscopes have a short working distance, making it difficult to see deep within large biological samples such as an intact brain. Slicing the tissue with a blade can reach deeper, but this often distorts or destroys the fine structures that scientists want to study. By embedding a sample in a light-sensitive hydrogel, Wang *et al*. demonstrated a gentler approach using a precise ray or sheet of light to dissolve or cut away tissue layer by layer. After each layer is removed, the newly exposed surface is imaged, allowing for a complete, high-resolution, three-dimensional reconstruction without damaging physical contact. —Stella M. Hurtley

Science p. 257, 10.1126/science.adr9109

CELL BIOLOGY

Observing neurotransmission in a flash

Neurons transmit signals through synaptic vesicle release, but the nanoscale dynamics of this process have been unclear. Tao *et al*. revealed these dynamics in hippocampal synapses using optogenetics and time-resolved cryo–electron tomography (see the Perspective by Lichter). Their innovative approach captures synaptic vesicle exocytosis at millisecond resolution, identifying the "kiss-shrink-run" pathway in which synaptic vesicles briefly contact the presynaptic membrane, shrink significantly, and then detach for rapid recycling. This mechanism unifies competing neurotransmitter release models and elucidates the underpinnings of synaptic efficiency and reliability. —Stella M. Hurtley

Science p. 258, 10.1126/science.ads7954; see also p. 236, 10.1126/science.aec0091

NEUROSCIENCE

Wild neuronal recordings

The neural mechanisms underlying complex navigation behaviors have largely been studied in small-arena laboratory experiments, which are very different from real-world environments. It is unknown whether the spatial cells discovered in the brain's navigation circuit are relevant for real-world movement. Palgi *et al*. recorded from single neurons in bats flying over an island (see the Perspective by Zhou and Knierim). They discovered that head-direction cells, neurons that encode an animal's orientation, represent direction in a global, compass-like manner, not as a mosaic of local representations. This means that a given neuron exhibits the same preferred compass direction throughout the geographical extent of the natural environment. Head-direction cells were stable despite the dynamics of the moon and the stars. These findings establish the relevance of head-direction cells to real-world navigation. —Peter Stern

Science p. 259, 10.1126/science.adw6202; see also p. 239, 10.1126/science.aeb6967

GLOBAL SEA LEVEL

An old story

Global mean sea level (GMSL) and climate are inextricably intertwined, a point of great relevance in our warming world. Clark *et al*. present a reconstruction of GMSL changes over the past 4.5 million years that accounts for temperature-driven changes in the oxygen isotopic compositions of the main ice sheets, as well as changes in ice volume and ocean temperature. This approach provides important insights about ice sheet volumes and variability and the forces that drive conditions over time. —Jesse Smith

Science p. 260, 10.1126/science.adv8389

ELECTROCHEMISTRY

Inorganic additives stabilize ionomers

Anion-exchange membrane water electrolyzers represent one of the most economical methods for producing green hydrogen. However, their

limited durability, particularly during operation with pure water, remains a major obstacle to widespread adoption. Hou *et al*. investigated ionomer binder design in greater depth and demonstrated that appropriately introduced inorganic additives can help to prevent oxidative degradation of the ionomer while preserving excellent mechanical integrity and ionic conductivity, ultimately extending device lifetime (see the Perspective by Kim). The enhanced stabilization originates from cross-linking between the additive and the ionomer and is broadly applicable across combinations of catalysts, additives, and ionomers. —Jack Huang

Science p. 294 10.1126/science.adw7100; see also p. 237, 10.1126/science.aeb7886

2D MATERIALS

Making stable trions

Semiconductors can host multiparticle electron-hole states, which are usually short-lived due to electron-hole recombination. A promising system for studying their equilibrium properties are double layers of transition metal dichalcogenides, in which the carrier density in each layer can be independently controlled. Qi *et al*. and Nguyen *et al*. engineered an equilibrium quantum liquid of trions consisting of, for example, two holes and an electron in Coulomb-coupled molybdenum and tungsten diselenide ($MoSe_2$/WSe_2) double layers (see the Perspective by Ouyang and Shi). To achieve this, the researchers tuned the ratio of electron density in the $MoSe_2$ layer to the hole density in the WSe_2 layer and used transport, capacitance, and optical measurements to probe the system. —Jelena Stajic

Science p. 299, 10.1126/science.adn4584, p. 304, 10.1126/adn5886; see also p. 241, 10.1126/science.aec0463

IMMUNOLOGY

Ensuring recombination fidelity

To ensure a diverse repertoire of antigen receptors in T and B cells, their genes undergo V(D)J recombination, which is initiated by the formation of DNA double-strand breaks by the RAG complex and followed by DNA repair through the nonhomologous end-joining pathway. Using a CRISPR/Cas9–based screen, Libri *et al*. identified the RNA/DNA helicase senataxin as playing a previously uncharacterized role in V(D)J recombination. In cells and in mice, senataxin facilitated the repair of RAG-induced double-strand breaks and limited abnormal end joining, thus ensuring the fidelity of recombination. —John Foley

Sci. Signal. (2025) 10.1126/scisignal.adv8801

KIDNEY DISEASE

Go low to target neutrophils

Crescentic glomerulonephritis (cGN) is an autoimmune disease treated with glucocorticoids, but the therapeutic mechanisms are not fully understood. Yin *et al*. used single-cell sequencing of immune cells in a mouse model of cGN and found a proinflammatory subset of neutrophils that was a potential target of glucocorticoid therapy. Depleting these neutrophils improved disease phenotypes and their adoptive transfer made them worse. Repeated low doses of glucocorticoids, without an initial high dose, prevented neutrophil differentiation in the kidneys. Patients with high disease activity had similar neutrophils in inflammatory niches, which were decreased in number after repetitive low-dose glucocorticoid treatment. These results show that proinflammatory neutrophils drive cGN progression and highlight how low-dose glucocorticoid treatment may be a therapeutic option to slow disease progression. —Brandon Berry

Sci. Transl. Med. (2025) 10.1126/scitranslmed.adu0351

DRUG MECHANISMS

Preeclampsia drug as a glioblastoma therapy

Since 1953, the vasodilator hydralazine has been a primary therapy for preeclampsia, pregnancy-induced hypertension, but its mechanism of action has been unknown. Shishikura *et al*. used a hydralazine analog to covalently capture and identify its target, 2-aminoethanethiol dioxygenase (ADO). This enzyme oxidizes proteins with N-terminal cysteines, marking them for proteasomal degradation. Inactivation of ADO protects its substrates, including proteins in calcium-signaling pathways that mediate vasodilation. Aggressive forms of glioblastoma express high levels of ADO, suggesting that hydralazine might interfere with the growth of this brain cancer. A single exposure efficiently induced senescence in glioblastoma cells, raising the possibility for use of ADO inhibitors in the treatment of glioblastoma. —Craig Cameron

Sci. Adv. (2025) 10.1126/sciadv.adx7687

RESEARCH ARTICLE SUMMARY

PROTEOMICS

Mapping early human blood cell differentiation using single-cell proteomics and transcriptomics

Benjamin Furtwängler, Nil Üresin, Sabrina Richter, Mikkel Bruhn Schuster, Despoina Barmpouri, Henrietta Holze, Anne Wenzel, Kirsten Grønbæk, Kim Theilgaard-Mönch, Fabian J. Theis*, Erwin M. Schoof*, Bo T. Porse*

Full article and list of author affiliations: https://doi.org/10.1126/science.adr8785

INTRODUCTION: Our ability to characterize complex biological systems at single-cell resolution has facilitated the in-depth resolution of cellular subsets, states, and differentiation trajectories. Given that such studies, to a large extent, have relied on single-cell RNA sequencing (scRNA-seq), it is currently unknown whether emerging technologies such as single-cell proteomics by mass spectrometry (scp-MS) could uncover additional systems information, which would have broad implications for our understanding of fundamental biological processes.

RATIONALE: Proteins are the primary executioners of genetic information and because there is a far-from-ideal correlation between mRNA and the proteins they encode, we hypothesized that scp-MS could complement scRNA-seq in the characterization of complex biological systems. In this work, we therefore generated a scp-MS dataset of the human hematopoietic stem and progenitor cell (HSPC) hierarchy and integrated it with a matched scRNA-seq dataset, allowing us to follow the two modalities across hematopoietic differentiation.

RESULTS: We performed scp-MS and scRNA-seq on the human CD34+ HSPC hierarchy and demonstrated that the proteome data resolved HSPC heterogeneity in accordance with accompanying flow cytometry data. Using a variational autoencoder deep learning tool, we integrated the transcriptomic and proteomic modalities resulting in a joint embedding. Trajectory analysis demonstrated that, unlike the individual modalities, the integrated dataset more accurately predicted differentiation end states, highlighting the complementarity of the two modalities and their combined advantage in the characterization of hematopoietic differentiation.

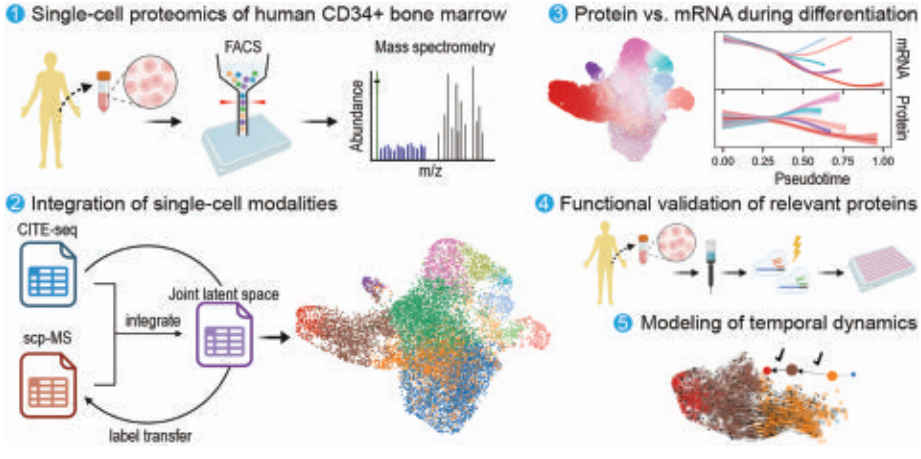
We also compared the concordance of mRNA and protein information across hematopoietic differentiation and found them to correlate well in more differentiated progenitor populations. By contrast, the two modalities correlated poorly in the most immature stem and progenitor subsets. We also identified proteins whose abundances were better explained by other proteins (such as protein complex partners) rather than their cognate mRNAs.

To showcase the ability of scp-MS to uncover proteins of functional importance in stem cells, we selected proteins whose abundances were enriched in stem cells at the protein but not mRNA level and functionally inactivated them using gene editing in human HSPCs. As examples, proof-of-concept functional analyses demonstrated that inactivation of TALDO1, a key enzyme in the pentose phosphate pathway, and H1F0, a histone 1 linker protein associated with quiescent cells, yielded in vitro phenotypes consistent with their functional roles in blood stem cells.

Finally, to model the temporal dynamics of mRNA and protein across differentiation, we developed single-cell protein velocity (scProtVelo), which explains observed mRNA and protein abundances across time through learned kinetic parameters. In contrast to corresponding analyses relying solely on mRNA, scProtVelo correctly resolved the information flow in several trajectories, further highlighting the power of combining mRNA and protein level data at the single-cell level.

CONCLUSION: We have demonstrated the potential of scp-MS both alone and in combination with scRNA-seq for uncovering biological insights even in an extensively studied system such as the human HSPC compartment. As scp-MS technology continues to develop with respect to sensitivity, throughput, and possibilities for assessing post translational modifications, our understanding of the proteome at single-cell resolution will continue to improve, with far-reaching consequences for biomedical research. □

*Corresponding author. Email: bo.porse@finsenlab.dk (B.T.P.); erws@dtu.dk (E.M.S.) fabian.theis@helmholtz-munich.de (F.J.T.) Cite this article as B. Furtwängler *et al.*, *Science* **390**, eadr8785 (2025). DOI:10.1126/science.adr8785



From single-cell proteomics (scp-MS) to multimodal analysis. (1) Application of scp-MS workflow and (2) subsequent integration of scRNA-seq and scp-MS data enable (3) the comparison of mRNA and protein abundance during differentiation at single-cell resolution. (4) Identification and validation of proteins of functional importance and (5) recovery of temporal dynamics through a translation model. [Figure partially created with BioRender.com]

PROTEOMICS

Mapping early human blood cell differentiation using single-cell proteomics and transcriptomics

Benjamin Furtwängler[1,2,3]†, Nil Üresin[1,2,3]†, Sabrina Richter[4,5]†, Mikkel Bruhn Schuster[1,2], Despoina Barmpouri[1,2], Henrietta Holze[3], Anne Wenzel[1,2], Kirsten Grønbæk[2,6,7], Kim Theilgaard-Mönch[1,2,6,7], Fabian J. Theis[4,5]*, Erwin M. Schoof[3]*, Bo T. Porse[1,2,6]*

Single-cell RNA sequencing (scRNA-seq) has facilitated the characterization of cell state heterogeneity and recapitulation of differentiation trajectories. However, the exclusive use of messenger RNA (mRNA) measurements comes at the risk of missing important biological information. We leveraged recent technological advances in single-cell proteomics by mass spectrometry (scp-MS) to generate an scp-MS dataset of an in vivo differentiation hierarchy encompassing >2500 human CD34+ hematopoietic stem and progenitor cells. Through integration with scRNA-seq, we identified proteins important for stem cell function, which were not indicated by their mRNA transcripts. Further, we showed that modeling translation dynamics can infer cell progression during differentiation and explain substantially more protein variation from mRNA than linear correlation. Our work offers a framework for single-cell multiomics studies across biological systems.

Single-cell RNA sequencing (scRNA-seq) of the human hematopoietic stem and progenitor cell (HSPC) compartment has changed our understanding of the hematopoietic system by revealing the gradual transition through transcriptome states during HSPC differentiation toward functional blood cells (*1*–*4*). Capturing these transitional cell states from single-cell snapshots has enabled in-depth characterization of cell state heterogeneity and recapitulation of differentiation trajectories (*5*, *6*). Furthermore, modeling RNA velocity from mRNA splicing kinetics using scRNA-seq data has enabled the study of temporal dynamics during differentiation (*7*–*9*), although its application in hematopoiesis has faced challenges as a result of time-dependent kinetic rates (*10*, *11*). Although very successful, these methods have mostly been based on the transcriptome while the actual workhorse for cellular function, the proteome, has not been widely accessible for single-cell analysis. The importance of proteome characterization has been shown in many multiomics studies, encompassing mRNA and protein measurements on the bulk population level by revealing that transcript abundance can be a poor proxy for protein abundance (*12*–*16*) and that the proteome can reveal more phenotypical information (*17*, *18*). Furthermore, these bulk multiomics datasets have enabled the modeling of translation kinetics during cell differentiation (*16*), revealing the temporal dynamics between the two modalities (*19*–*21*). On the single-cell level, protein measurements have mostly been limited to targeted antibody-based cell surface epitope assays such as cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq), which were nonetheless instrumental in resolving cell states inaccessible with mRNA measurements (*3*, *22*, *23*) and even enabling temporal modeling (*24*). Thus, direct, untargeted, and high-throughput measurements of proteins in single cells have the potential to further improve our understanding of biological systems by providing more phenotypical information about observed cell states and by revealing temporal dynamics in the system.

Recent technological advances have enabled measuring protein abundance in individual cells using single-cell proteomics by mass spectrometry (scp-MS) (*25*). Specifically, an early iteration of this workflow (*26*) uses sample multiplexing through isobaric tags enabling measuring 14 to 16 single cells in a single MS run. The addition of a carrier channel provides more peptide ion copies to the MS to facilitate peptide identification. Based on this method, we have previously introduced our 384-well plate–based scp-MS workflow (*27*, *28*) in which single cells are isolated through fluorescence-activated cell sorting (FACS), which enables the incorporation of antibody-based surface markers to isolate or annotate populations of interest.

Although scp-MS has reached a technological level that allows for the quantification of a considerable part of the proteome, reports describing its application to in vivo differentiation hierarchies remain scarce. This poses the questions of whether scp-MS can be upscaled for robust data generation of complex biological systems such as the human HSPC compartment and whether it captures phenotypical information that is complementary to mRNA-based approaches. This would make scp-MS well-suited for single-cell multiomics analyses, if integration of scRNA-seq and scp-MS data is additionally demonstrated.

We present a scp-MS dataset encompassing >2500 cells, which recapitulates the HSPC hierarchy through relative protein abundance changes. We subsequently computationally integrate scp-MS with scRNA-seq, generating a single-cell multiomics dataset. We discover proteins involved in stem cell function that are not apparent on the mRNA level and highlight processes during lineage specification that are better described on the protein level. Finally, we quantify temporal dynamics between the modalities and show that these can be modeled through a translation kinetics model (single-cell protein velocity), which we termed scProtVelo.

Results

A single-cell proteomics dataset of FACS isolated human HSPCs

We isolated CD34+ bone marrow (BM) HSPCs from six healthy donors and stained the cells with a surface marker panel, enabling the isolation of defined subpopulations through FACS (*2*) (Fig. 1A). With the aim of measuring protein abundance in single cells, we applied our scp-MS workflow (*27*, *28*) starting with the deposition of single cells into the wells of a 384-well plate while recording the FACS parameters for later analysis (Fig. 1B). We applied two different sorting strategies resulting in two kinds of multiplexed sets for scp-MS: First, we used the classical gating strategy to isolate specific subpopulations, and then distributed them evenly across the 14 single cells per multiplexed set ("enriched populations"). Secondly, we sorted any CD34+ cell resulting in the random sampling of the entire HSPC population ("total HSPC"). For the MS measurement of each multiplexed set, we used our recently described real-time search-assisted acquisition (RETICLE) (*28*). For downstream analysis, we applied our computational single-cell proteomics readout of expression (SCeptre) workflow (*27*), which expands on the functions of Scanpy (*29*) for the processing of scp-MS data, including cell filtering, batch correction, and data normalization (fig. S1, A to D; see also methods).

SCeptre processing yielded a dataset with more than 2900 proteins quantified across 2500 cells, albeit with variable data completeness (Fig. 1C) and, on average, 68% missing values per cell, which is comparable

[1]The Finsen Laboratory, Rigshospitalet, Copenhagen University Hospital, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark. [2]Biotech Research and Innovation Centre, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark. [3]Department of Biotechnology and Biomedicine, Technical University of Denmark, Lyngby, Denmark. [4]Helmholtz Zentrum München – German Research Center for Environmental Health, Institute of Computational Biology, Neuherberg, Germany. [5]TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany. [6]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark. [7]Department of Hematology, Rigshospitalet, Copenhagen University Hospital, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark. ***Corresponding author. Email: bo.porse@finsenlab.dk (B.T.P.); erws@dtu.dk (E.M.S.); fabian.theis@helmholtz-munich.de (F.J.T.)**
†These authors contributed equally to this work.

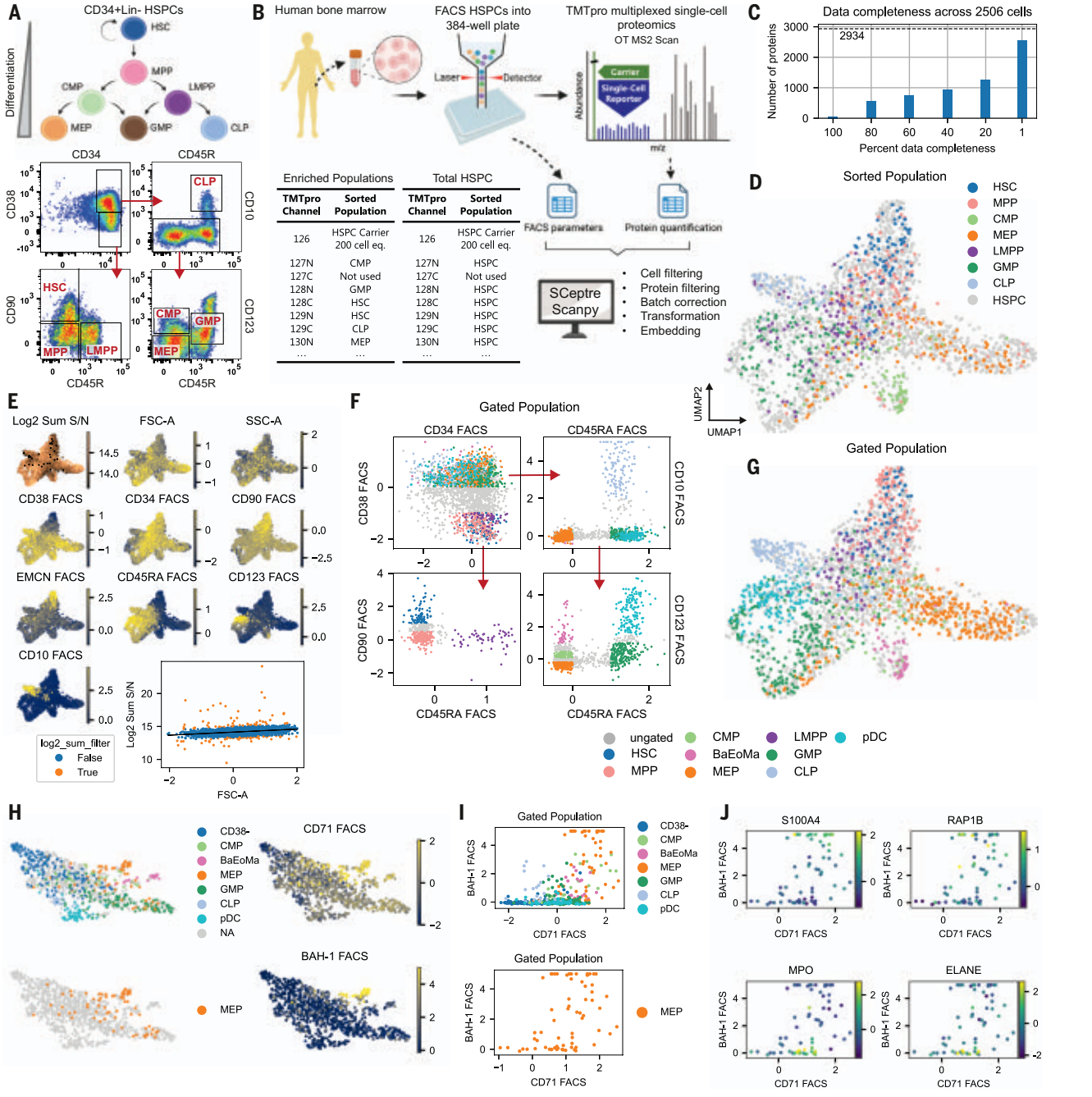


Fig. 1. A single-cell proteomics dataset of FACS-isolated human HSPCs. (**A**) The HSPC differentiation hierarchy (top) and sorting strategy (bottom). (**B**) Experimental workflow in which two different kinds of multiplexed sets were generated by sorting cells from specific gates ("enriched populations") or any Lin-CD34+ cell (total HSPC). (**C**) Data completeness of scp-MS dataset. (**D**) UMAP of cells from scp-MS data, annotated with their respective sorting gate. (**E**) Overlay of total MS signal and FACS parameters onto the UMAP. Scatterplot shows the relationship between FACS forward scatter area and scp-MS total signal. Cells were removed if the absolute deviation of the Log2 Sum S/N from the regression line was above 0.8. (**F**) Computational gating strategy for retrospective annotation of cells from both enriched populations and total HSPC sorts. (**G**) UMAP annotated with gated populations from (F). (**H**) UMAP of second scp-MS dataset with FACS markers CD71 and BAH-1 enabling positive selection of MEPs. (**I**) Scatterplot showing the putative true BAH-1+CD71+ MEP subset. Populations were annotated using the strategy in (F). Additional FACS markers reveal true MEPs. (**J**) Scatterplots from (I) with overlaid abundance levels of MEP (top) and GMP (bottom) marker proteins.

or even better than previous results (*30*). From the uniform manifold approximation and projection (UMAP) (*31*) embedding of the protein abundance data, we did not observe any batch effects based on the tandem mass tag (TMT) channel, individual donors, or plates (fig. S1, B to D), indicating that technical and biological (donor) variations were largely removed. We further validated this using principal component regression (*32*), which revealed that only very low fractions of the variance could be explained by MS run, TMT label, donor, or age (fig. S2A). Overlaying the cell identities derived from the two different cell-sorting strategies onto the UMAP embedding revealed clustering and branching of the sorted (enriched) populations based on the HSPC differentiation hierarchy (Fig. 1D). Hematopoietic stem cells (HSCs)

and multipotent progenitors (MPPs) were primarily located at the top of the hierarchy as a mixed and heterogeneous population, which has also been previously observed in scRNA-seq (*33*). Downstream of HSCs/MPPs followed lymphoid-primed multipotent progenitors (LMPPs), which tended to be located upstream of granulocyte-macrophage progenitor (GMP) and common lymphoid progenitor (CLP) branches. Some megakaryocyte-erythroid progenitors (MEPs) clustered together with GMPs (further investigated below). Moreover, we observed cells sorted as common myeloid progenitors (CMPs) clustered together with both myeloid branches but also forming a separate cluster that we later determined to be basophil, eosinophil, and mast cell progenitors (BaEoMa) (see below). This aligns well with findings that immunophenotypical CMPs constitute a heterogeneous population (*34*–*36*) and with functional studies demonstrating BaEoMa potential within the immunophenotypic CMP population (*37*–*39*). Taken together, these results showed that by SCeptre processing of the scp-MS data, we successfully removed batch effects from technical covariates without the need to define them.

Recording the FACS parameters during sorting enabled us to overlay that information onto the UMAP together with the total scp-MS protein signal (Fig. 1E). This demonstrated the alignment of FACS markers and the scp-MS UMAP embedding. Cells differed in their scatter properties, indicative of differences in cell size and granularity. The differences in forward-scatter area (FSC-A) correlated well with the total protein signal (Fig. 1E), indicating a higher protein content for larger cells. Furthermore, our data suggest that Endomucin (EMCN) might serve as an alternative surface marker for enriching long-term–HSCs (*40*) as commonly used markers CD90 and CD49f (fig. S2B) exhibited lower or no specificity in our dataset.

We subsequently used the recorded FACS parameters to perform retrospective computational gating (Fig. 1F). We refined the gating strategy in the CD38+CD10− fraction (lower right panel) by moving the CMP gate down to CD45RA-CD123dim to incorporate a CD45RA-CD123+ gate for BaEoMa (*39*). Furthermore, we moved the GMP gate to CD123−/dim to incorporate a CD45RA+CD123+ population of progenitors for dendritic cells (pDC) (*2*). Overlaying these annotations onto the proteomics-derived embedding revealed well-clustered populations in line with the HSPC differentiation hierarchy (Fig. 1G).

We observed that some MEPs clustered together with the GMP branch (Fig. 1, D and G). MEPs are purified based on the absence of specific markers and this population has been shown to be impure in functional assays (*34*, *41*). To improve the MEP annotation, we generated an additional scp-MS dataset with 922 cells, incorporating CD71 and CD110/BAH-1 FACS markers to further specify the erythroid and megakaryocyte progenitors within the conventional MEP gate (*34*) (Fig. 1H and fig. S3). The combination of CD71 and BAH-1 clearly delineated MEPs, allowing the separation of those MEPs that clustered with GMPs (Fig. 1, H and I). We found proteins associated with MEP differentiation, such as S100A4 and RAP1B, to be highly abundant in the BAH-1 positive MEPs whereas proteins associated with granulocytic differentiation, such as ELANE and MPO, to have high levels in some of the putative false MEPs with low BAH-1 levels (Fig. 1J). This observation corroborates the proximity of the BAH-1 negative cells in the MEP gate to the GMP branch. Overall, our scp-MS data recapitulated the human HSPC hierarchy and highlighted certain limitations of low-dimensional FACS marker combinations in resolving the cellular heterogeneity.

Unsupervised clustering reveals distinct cellular differentiation stages and functional protein covariance

Next, we used unsupervised clustering of the scp-MS data to group cells into 11 clusters (Fig. 2A) that were primarily named after their most prevalent population, based on FACS information and protein abundance. Note the switch from MEP to the more common annotation of erythroid-megakaryocyte progenitors (EMPs). Furthermore, we named the first two clusters downstream from HSCs Progenitors 1 and 2, as they should contain the early bifurcation of HSPCs into LMPPs and EMPs, which could not be resolved fully with this clustering approach. Next, we evaluated the quantitative accuracy of the scp-MS data by comparing fold changes generated from cluster averages (pseudobulks) to an external bulk proteomics dataset derived from human HSPCs (HSC, MEP, CMP, GMP) (*42*) (Fig. 2, B and C). We also included the computationally gated scp-MS populations from Fig. 1G, as they might better resemble the cell composition of the bulk-sorted populations. Pearson correlation values between protein log2 fold changes of scp-MS–based populations and the four bulk populations revealed that scp-MS populations clustered together by similarity and had the highest correlation coefficients with their corresponding bulk population. The GMP and CMP bulk populations had slightly lower and much lower correlations, respectively, likely due to minor differences between the bulk and single-cell FACS gating. Furthermore, we used the bulk data to benchmark cell-wise normalization based on total signal versus median-ratio normalization, and found the latter to perform slightly better (fig. S4).

To evaluate the purity and integrity of the scp-MS clusters, we overlaid cluster labels onto the computational gating scheme from Fig. 1F (Fig. 2D and fig. S5). Most clusters were well-restricted to their respective FACS gate. Only LMPPs, Progenitors 2, and to some degree, Progenitors 1, were distributed over multiple FACS gates, indicating that these populations contain a mixture of differentiation stages. This highlights the issues of using a primarily FACS-based approach for cell annotation in a cluster-based context and in the absence of a reference map. Therefore, the UMAP annotation in Fig. 2A should be considered as an intermediate toward the more refined annotation of the scp-MS data presented later. We also note that the EMP cluster should give rise to another branch containing megakaryocyte progenitors; however, we did not observe any putative branch, which might be due to their rarity in BM or their depletion through the inclusion of CD61 megakaryocyte marker in the lineage cocktail in FACS.

We next analyzed differential protein abundance to identify marker proteins for the 11 clusters (Fig. 2, E and F). For example, HSCs showed high levels of H1.0 linker histone (H1F0), indicating a compact chromatin structure and quiescent cell state (*43*). Erythroid differentiation was marked by high levels of carbonic anhydrase I (CA1) (*44*). BaEoMas were marked by CD123 surface expression and high protein levels of Proteoglycan 2 (PRG2) and Charcot-Leyden Crystal Galectin (CLC), components of eosinophil granules (*45*). Granulocyte-monocyte-dendritic cell progenitors (GMDPs) displayed high levels of AZU1, PRTN3, ELANE, CTSG, and MPO, which are found in azurophil granules, the first granules to be formed during neutrophil differentiation (*46*). Progenitors for monocytes and common dendritic cells (MDPs) branched off from the GMDP cluster, marked by high levels of LGALS1 and PLD4, which could also be observed in scRNA-seq data (*47*) (fig. S6). Furthermore, MDPs showed increased surface expression of CD123 and high levels of proteins such as lysozyme (LYZ) and HLA-DRB1 (*3*). The cluster CLP/pre-pDC that likely contained a mixture of common lymphoid progenitors (CLP) and progenitors for plasmacytoid dendritic cells (pre-pDCs) shared markers with MDPs and progenitors for B cells (pBs). This is in line with the similarity of dendritic cell progenitors derived from myeloid or lymphoid lineages (*48*), however, they could be differentiated from MDPs by the lower levels of LYZ (*3*). Finally, the cluster that contained pBs was marked by surface expression of CD10 and high protein levels of topoisomerase IIβ (TOB2B), which is essential for B cell development (*49*). Thus, although HSPC differentiation is a continuous process, protein and FACS data are in good concordance on distinct abundance changes during lineage commitment.

To investigate protein covariation across HSPC differentiation, we performed unsupervised clustering of the most variable protein abundance profiles across clusters and performed gene ontology term enrichment analysis (*50*) on each protein cluster (Fig. 2, G and H). Protein

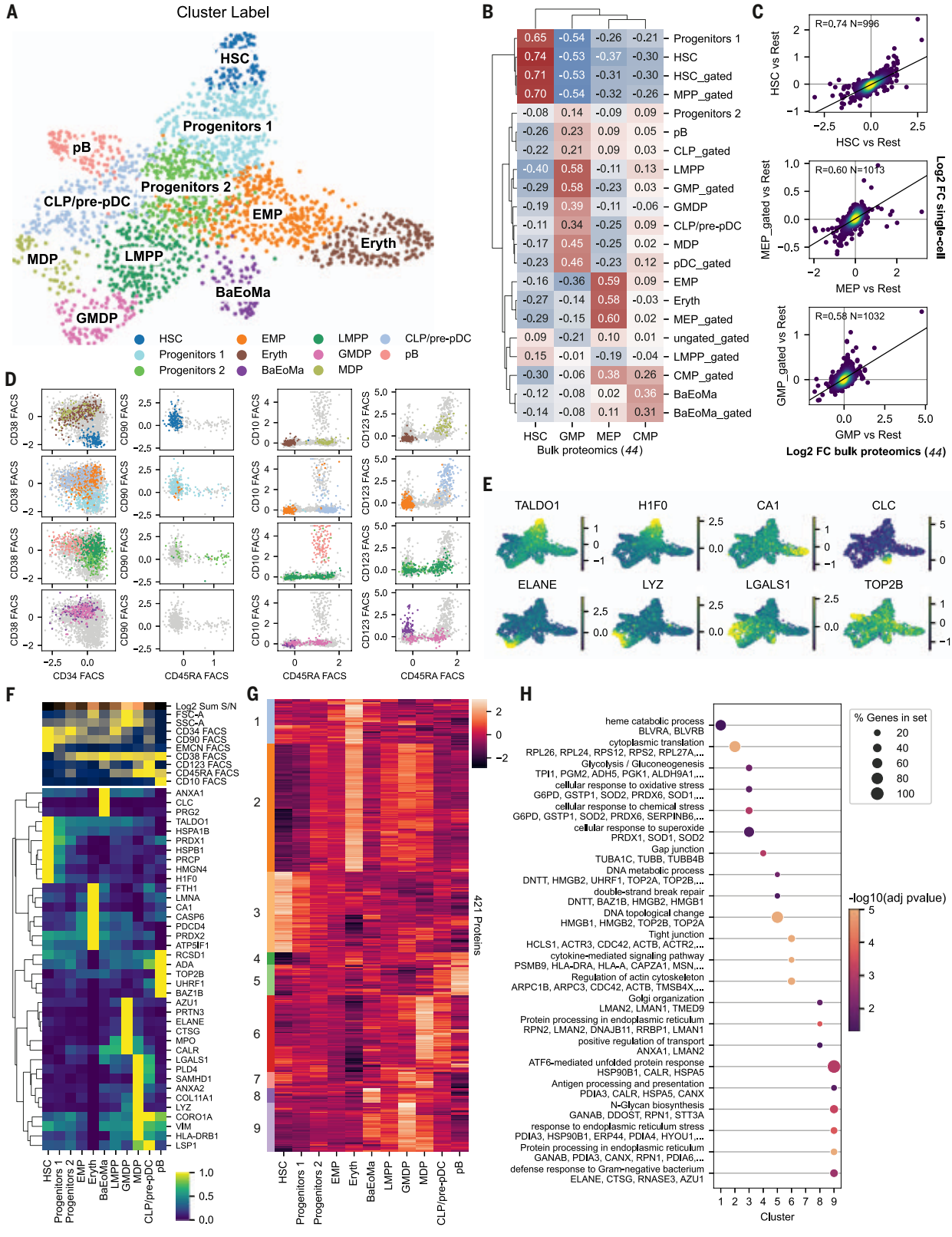


Fig. 2. Unsupervised clustering reveals distinct cellular differentiation stages and functional protein covariance. (**A**) UMAP of cells annotated with Leiden clusters. (**B**) Clustered heatmap showing the correlation of abundance changes between scp-MS measurements and bulk proteomics data (*42*), computed as Pearson correlation values of mean protein log2 fold changes of a population versus all other populations. (**C**) Scatterplots of selected comparisons from (B). (**D**) Cell clusters are visualized on the FACS gating scheme from Fig. 1F. Clusters were distributed over four panels for visibility. (**E**) UMAP annotated with the abundance of selected proteins. (**F**) Clustered heatmap of selected cluster-specific proteins. Values were mean-aggregated and normalized between 0 and 1. (Top) Aggregated FACS parameters and total signal. (**G**) Clustered heatmap of protein abundance across Leiden clusters. Values were mean-aggregated. (**H**) Dot plot of top enriched terms per protein cluster. −Log10(adj. *P*-value) was truncated to max. 5.

cluster 3 was highly abundant in HSCs and showed enrichment in glycolysis, likely reflecting their reliance on glycolysis instead of oxidative phosphorylation (*51*). Furthermore, HSCs showed high levels of proteins related to the response to oxidative and chemical stress, suggesting that these proteins are used to remove such stressors. The enriched terms found in other protein clusters related to the clusters' high abundance in certain populations (see supplementary text). Overall, our observations captured a diverse range of biological processes, adding further validity to the biological relevance of our dataset and highlighting the potential to analyze protein covariation at scale using scp-MS data.

Single-cell multiomics analysis of transcriptomics and proteomics of the human HSPC compartment

With the aim of performing an integrated analysis of mRNA and protein abundance in the HSPC compartment, we generated a CITE-seq (*22*) dataset of human HSPCs (Fig. 3A). We integrated data from four healthy adult humans using totalVI (*52*), resulting in a dataset of 9086 cells. From a UMAP embedding of the mRNA expression data, we did not observe any batch effects based on specific individuals (fig. S7A), indicating that technical and biological variation attributable to any specific donor sample were largely removed. We further validated this using principal component regression (*32*) which revealed that only very low fractions of the variance could be explained by donor or age, especially in the batch corrected, single-cell variational inference (scVI)-modeled (*53*) mRNA layer (fig. S7B). We annotated cells based on mRNA expression with a reference mapping from Azimuth (*4*) and, similar to the scp-MS data, subsetted GMPs into GMDP and MDP (fig. S7C). The cell annotation, in combination with the UMAP embedding, partition-based graph abstraction (PAGA) (*54*) connectivity, and surface marker expression indicated a good recapitulation of the human HSPC hierarchy (Fig. 3A and fig. S7, D and E). Furthermore, while cell cycle phase prediction resulted in an expected phase distribution across differentiation stages (fig. S7F), it also revealed strong cell cycle–dependent clustering of the UMAP embedding, which likely explains the unexpected connectivity between Early Eryth and LMPPs in the same phase. We also observed a very wide range of total mRNA counts and detected genes per cell differentiation stage, with HSCs having, on average, 3.45 times fewer counts and 2.1 times fewer detected genes compared with Late Eryth, in line with the low transcriptional activity in HSCs (*55*) (fig. S7G). Finally, investigating mRNA expression changes across populations revealed known markers (fig. S8). Since totalVI returns modeled and denoised mRNA expression values, similar to scVI (*53*), we benchmarked the fold change accuracy using previously published bulk mRNA data (*42*) and found that scVI-modeled and median-ratio normalized mRNA data showed higher correlation with the bulk data compared with the count-based data and similar centering around zero as the bulk data (fig. S9). Nonetheless, we note that the scVI-modeled values should be used with caution as they represent only the mean of a modeled distribution with potentially high uncertainty.

We integrated the scp-MS and CITE-seq datasets with graph-linked unified embedding (GLUE) (*56*) (Fig. 3B), a variational autoencoder–based tool designed for unpaired single-cell multiomics data integration. The resulting joint latent space successfully integrated "mRNA" cells from the CITE-seq dataset and "protein" cells from the scp-MS dataset, with a silhouette score of 0.03. Figure 3C shows the UMAP embedding of the resulting joint latent space with "mRNA" cell labels transferred to the "protein" cells by means of nearest neighbor matching. Biological variation from the individual modalities was preserved in the integrated embedding as metrics of cell type separation (*32*) were comparable before and after integration for both modalities (Fig. 3D and fig. S10A). Moreover, we found that many protein-level markers agreed with mRNA expression (fig. S10B). Overlaying the new labels onto the protein UMAP revealed a more nuanced clustering and annotation compared with the initial annotation (compare Fig. 3E and Fig. 2A). Further, we compared the cell identity of "protein" cells based on the scp-MS clusters to the transferred labels and indeed found that labels were not transferred across lineage branches (Fig. 3F). Furthermore, "protein" cells in scp-MS clusters Progenitors 1 and 2 were assigned, among others, EMP and LMPP labels, which suggested that the model learned to separate these lineages at an early stage based on protein abundance. Indeed, in these clusters we found early differences in expression of surface marker CD45RA (FACS) expression and scp-MS-derived proteins such as GAPDH, LCP1, and PSME1, which had higher levels in LMPP and its descendants, whereas the proteins SOD2 and ALDH1A1 had higher levels in EMPs (Fig. 3G and fig. S11). Moreover, the Azimuth label pre-mDC was transferred onto "protein" cells contained in the MDP cluster. Indeed, these cells could be separated from MDPs and pre-pDCs by distinctive protein abundance patterns of, for instance, CD123 (FACS) and scp-MS-derived proteins such as S100A11, COTL1, and especially LYZ (Fig. 3H). Taken together, these analyses demonstrated the successful integration of mRNA and scp-MS data as well as the resulting ability to transfer mRNA-derived labels onto "protein" cells, which facilitated the refinement of cell states in scp-MS data.

Trajectory analysis on joint latent space recapitulates HSPC differentiation

We found that the joint latent space represented the biological latent cell states, and we then asked whether we could identify differentiation trajectories and relevant driving factors. We therefore used the integrated latent space as input for a trajectory analysis using cellrank (*5*, *6*). We used the pseudotime kernel (Fig. 4A) that infuses directionality into a k-nearest neighbors graph to compute a transition matrix. The resulting terminal states and the cell-specific fate probabilities indeed represented the expected differentiation trajectories in our HSPC dataset (Fig. 4A, fig. S12, A to C, and supplementary text).

We subsequently investigated whether we could have obtained the same results on either one of the modalities independently. We found that the terminal states identified from the joint latent space were not readily recovered from either the CITE-seq or the scp-MS datasets only (Fig. 4B), even when more macrostates were considered (Fig. 4C). Furthermore, we compared the fate probabilities across the three models, which revealed a better separation of the pre-pDC and pre-mDC trajectories in the joint latent space compared with the individual modalities, specifically correct lineage assignments of CLP, pre-pDC, MDP, and pre-mDC improved from 86 to 91% on the RNA cells and from 65 to 95% on the protein cells (fig. S12D). Taken together, our results indicated that the cellrank model based on the joint latent space outperformed the models based on CITE-seq or scp-MS alone. These findings confirm our hypothesis that the two modalities provide complementary information about hematopoietic differentiation.

Subsequently, we used the trajectory analysis to investigate mRNA and protein abundance changes during differentiation. First, we analyzed HSC quiescence and differentiation by correlating mRNA or protein abundance to pseudotime in an early subset of cells (Fig. 4, D and E). This yielded a correlation-based ranking of mRNAs or proteins, with positive and negative correlations reflecting increases and decreases during differentiation, respectively. Secondly, we analyzed the process of lineage specification by correlating mRNA or protein abundance to fate probability for a particular lineage (Fig. 4D). This yielded a correlation-based ranking of mRNA or protein, with positive correlations reflecting a coordinated increase of abundance across single cells during differentiation into that lineage compared with all other lineages. Importantly, the cell feature vectors pseudotime and fate probability were derived from the joint latent space and thus enabled the comparison of mRNA and protein abundance by comparing the correlation vectors between modalities. We envisioned this analysis to be suitable for biological

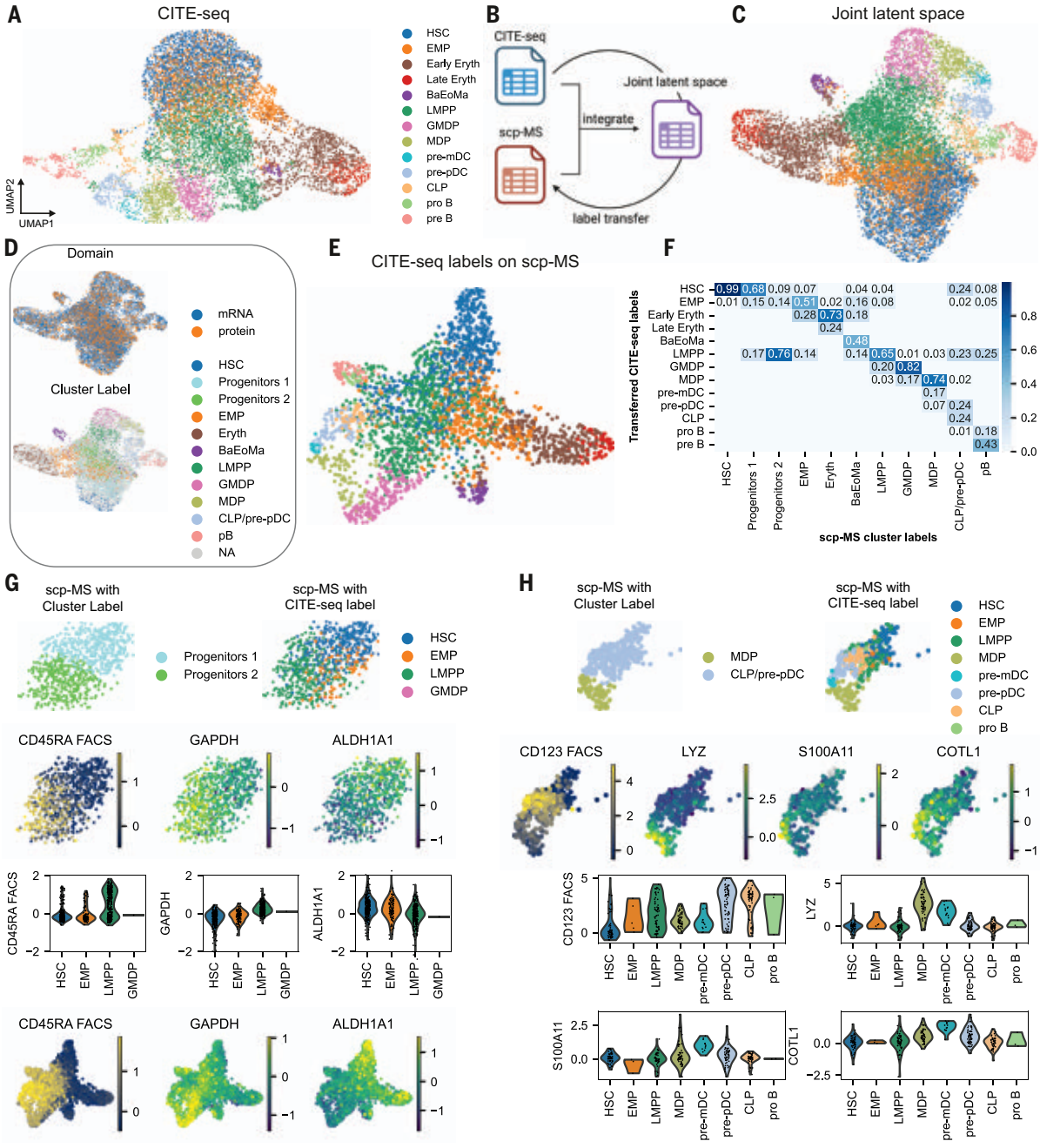


Fig. 3. Single-cell multiomics analysis of transcriptomics and proteomics of the human HSPC compartment. (**A**) UMAP embedding of annotated CITE-seq dataset. (**B**) Schematic of GLUE integration approach. (**C**) UMAP embedding of integrated dataset of CITE-seq and scp-MS, overlaid with CITE-seq cell annotation that was transferred to the "protein" cells. (**D**) Distribution of "mRNA" cells and "protein" cells (top). Overlaid cluster labels from scp-MS (bottom). (**E**) scp-MS UMAP embedding with transferred CITE-seq cell annotation. (**F**) Heatmap related to the label transfer from "mRNA" to "protein" cells. Rows are the cell annotations from CITE-seq and columns are the cluster labels from scp-MS. The values represent the fraction of cells in the respective scp-MS cluster that was assigned the respective CITE-seq label; thus, columns add up to one. (**G**) Validation of cell label transfer of the CITE-seq labels EMP and LMPP onto scp-MS labels Progenitors 1 and 2. In the first three rows, cells were subsetted to Progenitors 1 and 2. Surface marker and protein abundance profiles align with the CITE-seq label, revealing the early bifurcation into EMP and LMPP. (**H**) Validation of cell label transfer of the CITE-seq label pre-mDC onto scp-MS label MDP. Cells were subsetted to MDP and CLP/pre-pDC. Surface marker and protein abundance profiles align with the placement of the pre-mDC CITE-seq label, revealing an additional population missed by the scp-MS clustering.

systems characterized by gradual transitions as in HSPCs, in contrast to a cluster-based analysis in which fold changes between pseudobulk clusters are compared between the mRNA and protein levels. The correlation-based analysis does not require clustering and is computed on the single-cell level, emphasizing features that are changing along the axis of interest (pseudotime or fate probability). Additionally, with the fate probabilities from cellrank, it makes full use of the soft lineage assignments per cell, emphasizing lineage-specific changes.

To show the validity of this approach, we first analyzed only the mRNA expression using the "mRNA" cells. The analysis of HSC quiescence and differentiation on the early subset of cells revealed genes associated with HSCs, such as the transcription factor *ATF3*, which has been shown to

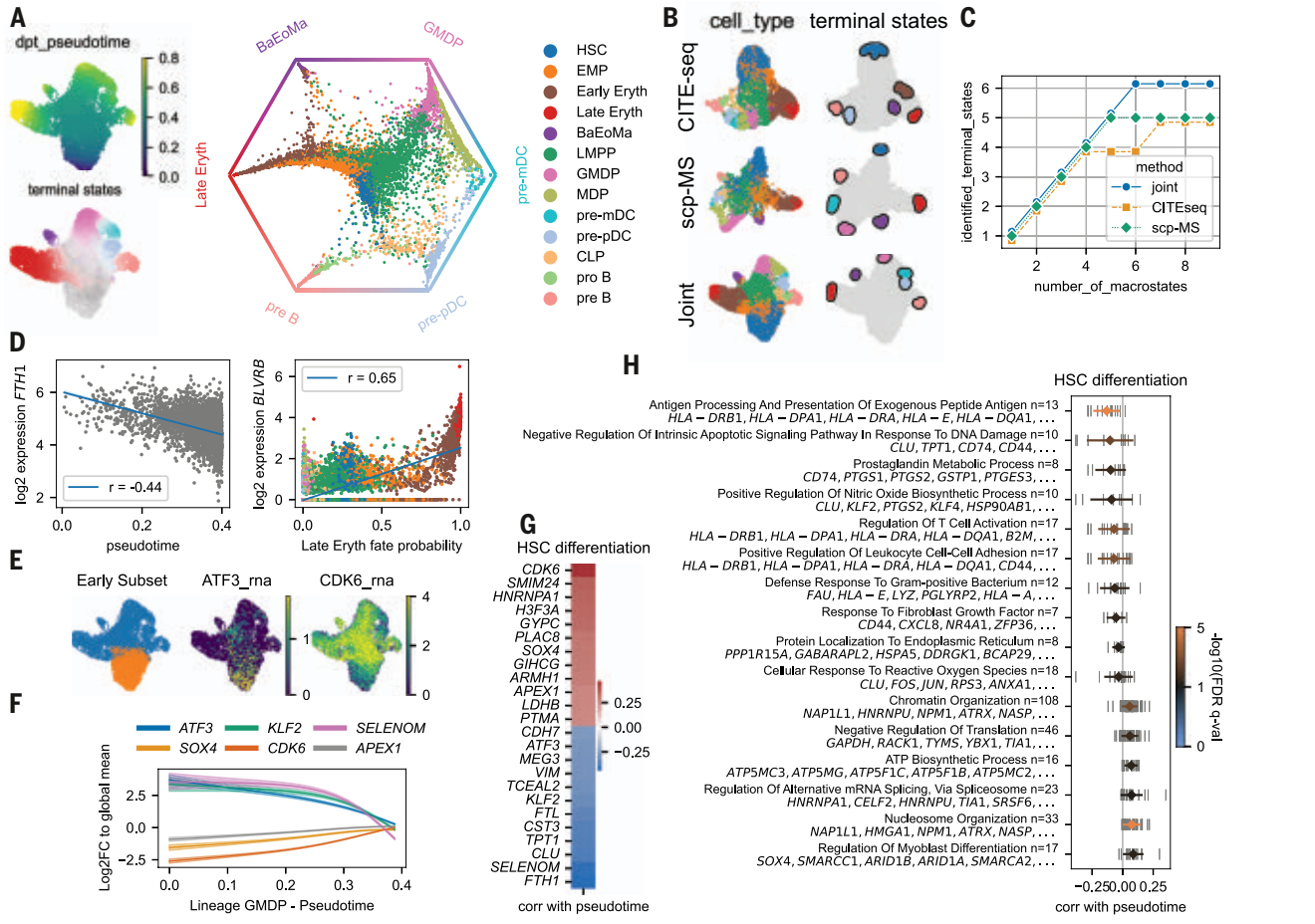


Fig. 4. Trajectory analysis on joint latent space recapitulates HSPC differentiation. (**A**) UMAP of joint dataset with overlaid pseudotime and terminal states determined by cellrank (left). Cellrank circular projection showing cells ordered by their relative fate probability for each lineage (right). (**B** and **C**) Comparison of cellrank terminal state detection between CITE-seq, scp-MS, and the joint dataset. (B) shows the top six detected terminal states and (C) shows the number of correct terminal states (*y*-axis) at different numbers of macrostates used for terminal state determination (*x*-axis). (**D**) Plots exemplify the calculation of correlation of gene expression with pseudotime (left) or fate probability (right). (**E** to **G**) mRNA expression analysis in an early subset of cells to investigate HSC differentiation. (E) Cells that were selected as early subset and mRNA expression of selected genes with high correlations. (F) Modeled gene expression profiles along pseudotime. The early part of the GMDP lineage was used for visualization. (G) Genes with the highest correlation or anticorrelation with pseudotime. (**H**) GSEA of HSC differentiation; *x*-axis shows correlation of expression with pseudotime. Vertical gray lines represent individual genes in the set. The horizontal bar represents the mean and standard deviation.

prevent HSC exhaustion (*57*), or *KLF2*, which is important for embryonic stem cell self-renewal (*58*) (Fig. 4, E to G). By contrast, genes associated with stem cell differentiation were *CDK6*, a key regulator of exit from quiescence (*59*), the transcription factor *SOX4* (*60*), and the endonuclease *APEX1* (*61*), which are both important for progenitor formation. Additionally, we modeled these gene expression profiles along pseudotime to enable a visualization (see materials and methods and Fig. 4F). Subsequently, we used the correlation vector to perform a pre-ranked gene set enrichment analysis (GSEA) (*50*, *62*) (Fig. 4H and table S1). This revealed known processes for HSC maintenance such as antigen presentation through major histocompatibility complex II (*63*) (*HLA-DRB1, HLA-DPA1, and HLA-DRA*), prostaglandin metabolism (*64*), nitric oxide metabolism (*65*), and reduction of reactive oxygen species. Processes enriched during HSC differentiation included nucleosome organization and increased expression of the ATP synthase, required for the switch from glycolysis to oxidative phosphorylation (*51*). Furthermore, the analysis of lineage specification revealed lineage-specific gene expression cascades and enrichment of relevant gene sets (figs. S13 to S17 and table S1). Taken together, these results indicated that the pseudotime represented the main axis of early stem cell differentiation and that cellrank recapitulated correct differentiation trajectories of the HSPC hierarchy and supported the validity of our correlation-based analysis.

Protein level information reveals additional insights into hematopoietic stem cell quiescence and differentiation

Subsequently, we continued with the analysis of protein abundance and directly compared the results with mRNA, by comparing correlation vectors between both modalities. For mRNA, we used the above calculated vectors based on counts but also included correlation vectors based on the modeled and denoised scVI data. For proteins, we repeated the analysis on the "protein" cells. The analysis of HSC quiescence and differentiation on the early subset of cells recapitulated the HSC protein markers we observed in Fig. 2F, whereas proteins associated with HSC differentiation were mainly related to DNA replication. Notably, on the mRNA level many of these genes showed only low correlations or even disagreements in the direction of change in abundance (Fig. 5A). Comparison between the full correlation vectors on the mRNA and protein levels revealed an overall weak correlation below 0.25 between these vectors (Fig. 5B). Moreover, this difference

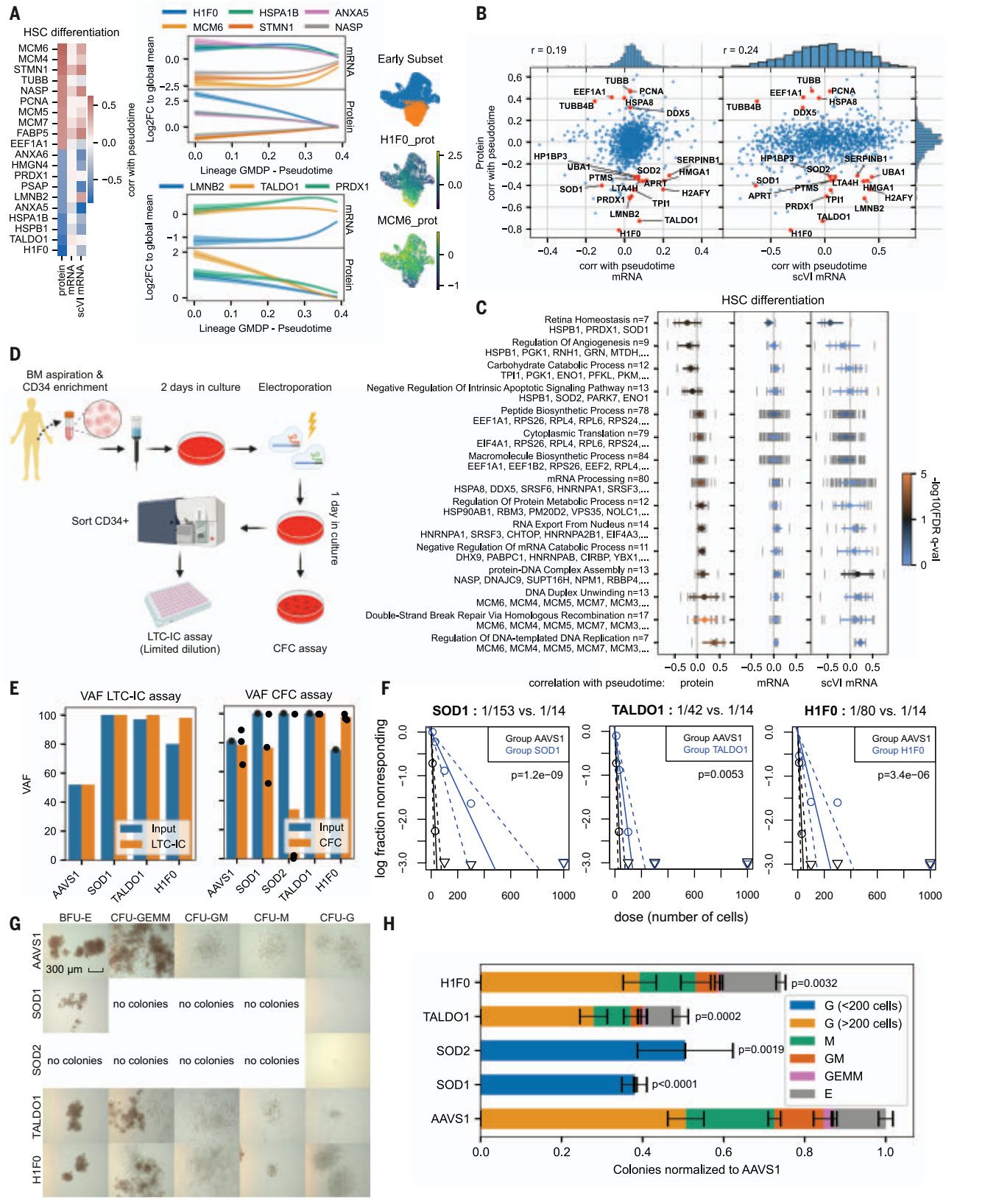


Fig. 5. Hematopoietic stem cell quiescence and differentiation are described differently on the protein level. (**A**) Protein abundance change analysis in an early subset of cells to investigate HSC differentiation. (Left) Proteins with the highest correlation and anticorrelation with pseudotime, including mRNA level information. (Middle, top) Modeled protein and mRNA abundance changes of selected genes. GMDP lineage was used as an example. (Middle, bottom) Modeled protein and mRNA abundance changes of selected genes that are not HSC-specific on the mRNA level. (Right) Cells that were selected as early subset and abundances of selected proteins. (**B**) Scatterplot comparing the correlation vectors between protein and mRNA (count-based and scVI-modeled). Selected genes were annotated. (**C**) GSEA analysis of correlation vectors of HSC differentiation.

Vertical gray lines represent individual genes in the set. The horizontal bar represents the mean and standard deviation. (**D**) Experimental workflow for the functional assessment of selected proteins. (**E**) Variant allele frequency (VAF) of CRISPR knockout. (**F**) Quantification of LTC-IC limiting dilution assay from one donor. Each data point was derived from seeding a given dilution in 10 wells. Statistical significance for the limited dilution experiments was assessed as described in (*71*). (**G**) Representative images of colonies of colony forming cell (CFC) assay. BFU, burst forming unit; CFU, colony forming unit; E, erythroid, GEMM, granulocyte/erythroid/megakaryocyte/macrophage; GM, granulocyte/macrophage; M, macrophage; G, granulocyte. (**H**) Quantification of CFCs. Total colony numbers were normalized to AAVS1 and statistical significance (total CFC numbers; unpaired *t*-test) for the individual knockouts (three technical replicates) was determined by comparison to their respective AAVS1 controls. Error bars represent standard deviations. Experiments in (E) to (H) were performed twice.

in the correlation vectors also resulted in the enrichment of gene sets that were not enriched on the mRNA level (Fig. 5C and table S1). For example, we observed an enrichment of proteins involved in carbohydrate metabolism (glycolysis) in HSCs, which were not enriched on the mRNA level, in line with typically low correlations of key metabolic pathways (*14*). Moreover, we found many proteins involved in the protection of HSCs from oxidative stress, including superoxide dismutases SOD1 and SOD2, Peroxiredoxin 1 (PRDX1), and TALDO1, a key enzyme of the pentose phosphate pathway that generates NADPH, which fuels the regeneration of reduced glutathione, a major cellular antioxidant (*66*). Apart from SOD1, a clear decrease of the levels of these proteins during early HSC differentiation could only be detected through scp-MS and not scRNA-seq (Fig. 5B). We also found many proteins involved in chromatin structure among HSC-correlating factors, including the previously mentioned H1 linker histone H1F0, the H1-like protein HP1BP3 required for HSC self-renewal (*67*); the histone macroH2A1 (H2AFY), associated with HSC homeostasis (*68*); and the chromatin regulator HMGA1, important for HSC regenerative capacity (*69*). Again, the decrease of these chromatin regulators during early HSC differentiation was better described through scp-MS. Finally, we observed strong differences between mRNA and protein levels of nuclear lamin B2 (LMNB2). It has been shown that Lamin-A:B stoichiometry regulates nuclear stiffness (*70*) and thus scp-MS could potentially better provide more information on nuclear morphology in single cells than scRNA-seq.

Finally, investigation of the proteins and their gene sets related to HSC differentiation revealed processes that were better described through scp-MS; for example, the increase in proteins involved in mRNA processing, RNA export from the nucleus, translation, and DNA replication (Fig. 5C and table S1). Notably, we investigated an early HSC cell subset of the data (Fig. 5A), in which detecting changes can be challenging as a result of the small amounts of proteins and especially mRNA in these cells and also due to the small effect sizes. This does not exclude the likely possibility that these terms would be enriched on mRNA level when comparing HSCs with more differentiated cells, in which effect sizes would be greater. Furthermore, we focused on proteins that showed absolute correlations above 0.2, as these proteins represented the potential additional information that could be gained from the scp-MS data. It should be noted that the mRNA data also contained distinct findings, which could be relevant to follow up on.

To validate the functional importance of selected proteins associated with the HSC state that have previously not been interrogated in humans, we performed CRISPR/Cas9-mediated knockout of SOD1, TALDO1, and H1F0 in human CD34+ BM cells (Fig. 5D). Subsequently, we conducted a long-term culture-initiating cell assay (LTC-IC) as a surrogate for HSC activity upon knockout (Fig. 5, D to F) as well as colony forming unit (CFU) and 3-week stromal co-culture assays to further assess the functional consequence of deleting these factors. LTC-IC limiting dilution analysis (*71*) revealed a reduction in LTC-IC frequency of TALDO1 and H1F0 deficient cells (1 in 42 and 1 in 80, respectively) compared with the AAVS1 control (1 in 14), suggesting a decreased ability to sustain long-term hematopoiesis upon knockout of these factors (Fig. 5F). Loss of TALDO1 and H1F0 also resulted in a mild decrease in the number of CFUs; however, the relative proportions of distinct colony types remained unaltered (Fig. 5, G and H), strongly suggesting a specific impact on the most immature cells, namely HSCs. This was further corroborated by stromal cell co-culture assays in which loss of these factors only had minor effects on cell proliferation (fig. S18, A and B), as well as on the frequencies of CD34+HSPCs and different myeloid cell populations (fig. S19). By contrast, SOD1-deficient HSPCs (a cytosolic superoxide dismutase) were largely unable to form colonies (only yielding small CFU-Gs, Fig. 5, G and H). This phenotype was associated with early proliferative cessation (fig. S18, A and B), a block especially in granulocytic differentiation in the stromal co-culture assay (fig. S19) and a major reduction in LTC-IC frequency (1 in 153; Fig. 5F). However, since SOD1 knockout HSPCs were nearly unable to form colonies, we could not determine the extent to which the latter is due to a specific HSC phenotype. Finally, we tested the functional consequences of knocking out SOD2 (a distant mitochondrial homolog of SOD1, 25% identify), which nearly replicated the SOD1 KO phenotype (Fig. 5, G and H, and figs. S18 and S19) except for the fact that SOD2-deficient HSPCs did not display a proliferative phenotype. Combined, these findings not only demonstrate the importance of SOD1 (and SOD2) for immature human HSPCs, but also highlight specific differences between mice and humans, as Sod1 and Sod2 null mice are not associated with any prominent HSPC phenotypes (*72*, *73*). In summary, the protein abundance trends during early HSC differentiation revealed additional information compared with scRNA-seq measurements and identified proteins and related pathways important for HSC function.

Proteomics-based trajectory analysis reveals protein level-specific functional covariation

Besides maintenance of the HSC state, we asked which protein-level programs are specific to particular lineages. For this, we analyzed lineage specification on the protein level in the integrated dataset and compared it directly with the mRNA level (Fig. 6, A to G, and figs. S20 to S24). This revealed many of the protein markers from Fig. 2F but also new proteins. For example, the Late Erythroid trajectory was characterized by the retained high abundance of ALDH1A1 (Fig. 6, A and C), as described above. Furthermore, we detected the transient abundance increase of markers for megakaryocyte progenitors such as RAP1B (*4*) and subsequent increase in levels of erythroid related proteins such as S100A4, PRDX2, and CASP6 (Fig. 6, A and C). Comparison between the full correlation vectors on the mRNA and protein levels for this lineage (Fig. 6D) revealed a much higher agreement of up to 0.7 compared with the HSC differentiation ($r = 0.19$ to 0.24). Nonetheless, differences between mRNA and protein correlation vectors propagated to differences in the GSEA, including processes involving ribosomal proteins (fig. S20A). In the GMDP lineage (Fig. 6, B, C, and E), we found gene sets related to protein glycosylation and endocytosis to be enriched only on the protein level (fig. S20B and table S1). The latter set contained Beta-2 microglobulin (B2M), which in our dataset was not well-correlated between mRNA and proteins as, for example, scp-MS measured an increase of B2M in the GMDP lineage whereas the mRNA measurement indicated lower relative expression (Fig. 6, E and F). Nonetheless, the B2M protein abundance profile corresponded well to its complex partners, HLA-A and HLA-B, in the MHC class I (Fig. 6, F and G).

Based on these results, we hypothesized that proteins having low concordance with their respective mRNA template still covaried with functionally related proteins. To investigate this, we performed a joint mRNA–protein covariation analysis. Since the trajectory analysis yielded six different lineages, for each mRNA and protein molecule the six correlations with lineage fate probabilities resulted in a vector of six

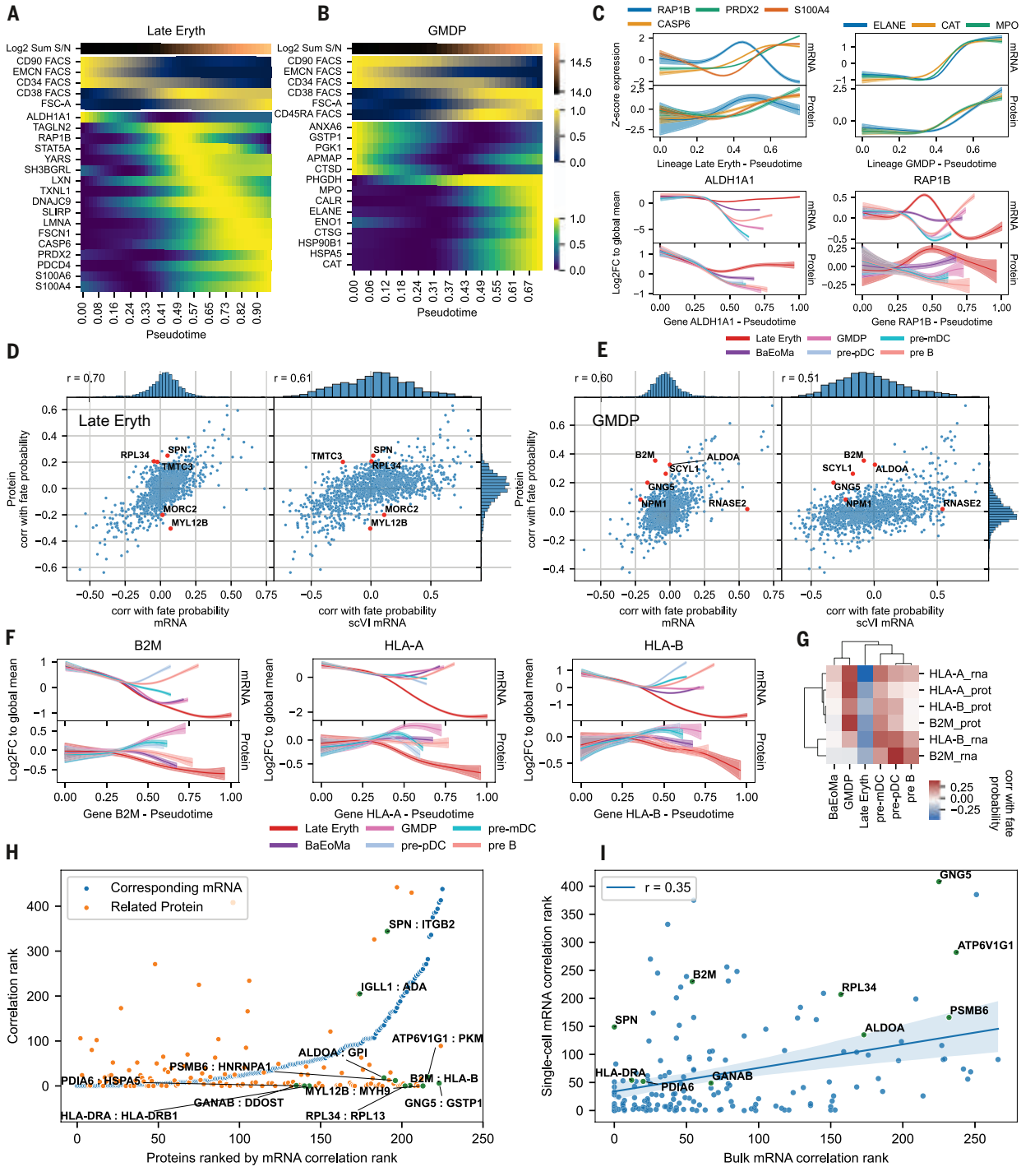


Fig. 6. Protein level trajectory analysis reveals protein level-specific functional covariation. (**A** and **B**) Heatmap showing abundance profiles of selected proteins in the Late Eryth (A) and GMDP (B) trajectories. (**C**) Modeled abundance profiles for protein and mRNA level of selected genes. (**D** and **E**) Scatterplot comparing the correlation vectors between mRNA and protein level in the Late Eryth (D) and GMDP (E) trajectories. (**F**) Modeled abundance profiles of MHC I complex members B2M, HLA-A, and HLA-B. (**G**) Clustered heatmap of abundance profiles of MHC complex members on both mRNA and protein level. (**H**) Scatterplot of proteins (*x*-axis) and the correlation ranks of both mRNA and protein molecules (*y*-axis). Proteins (*x*-axis) were sorted by the correlation rank of the corresponding mRNA (blue dots). Orange dots depict the related protein (same KEGG pathway) with the lowest correlation rank. (**I**) Comparison of mRNA correlation rank between bulk and scp-MS data.

values, representing the relative abundance across lineages. We focused on the 226 most variable proteins and their respective mRNA transcripts and calculated a correlation matrix between those 452 molecules. Then, for each protein we ranked all molecules by correlation and calculated the rank of the corresponding mRNA. We found that for a large fraction of proteins, the respective mRNA molecule ranked very high, indicating that it was among the best predictors for the protein abundance profile across lineages (Fig. 6H, blue dots). Subsequently, for each protein, we calculated the rank of the highest-ranking protein that was part of the same Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway. This revealed that many protein abundance profiles could be well-explained by a functionally related protein

(Fig. 6H, orange dots). Furthermore, looking at the proteins with low mRNA concordance and thus high mRNA ranks (right side of the plot), we indeed found that many protein abundance profiles could be better explained by functionally related proteins than their respective mRNA (Fig. 6H). Apart from the B2M-HLA-B pair, more examples include ALDOA and GPI, both part of glycolysis; MYL12B and MYH9, both part of the myosin complex; HLA-DRA and HLA-DRB1, both part of the MHC class 2 complex; and finally, protein pairs involved in protein processing in the endoplasmic reticulum, GANAB:DDOST and PDIA6:HSPA5 (Fig. 6H). Finally, to confirm that the observed low concordances between mRNA and protein was not purely due to dataset-specific measurement errors, we compared our results to an external bulk dataset with mRNA and protein measurements of 59 breast cancer cell lines (*74*–*76*). Calculating the mRNA correlation rank for each protein in this dataset revealed a positive correlation of 0.35 with our ranks (Fig. 6I). These results supported a general trend for some genes to exhibit low correlations between the mRNA and protein (*13*, *14*); however, this also indicates that there might be effects specific to the biological system or measurement technique.

scProtVelo models translation dynamics during differentiation

Finally, we hypothesized measuring both mRNA and protein expression changes during HSPC differentiation could enable the study of temporal dynamics in that system, namely transcription, translation, and degradation of mRNA and protein. Conceptually similar to how the time delay between unspliced and spliced transcript abundance changes enables RNA velocity (*7*–*9*) to infer splicing kinetics, the temporal delay between mRNA expression changes and the corresponding protein response encodes information about its translation dynamics (Fig. 7A).

Toward this endeavor, we developed scProtVelo, a latent variable model that explains observed mRNA and protein abundance changes over time through learned gene-specific transcription, translation, mRNA, and protein degradation rates. To validate the ability of scProtVelo to

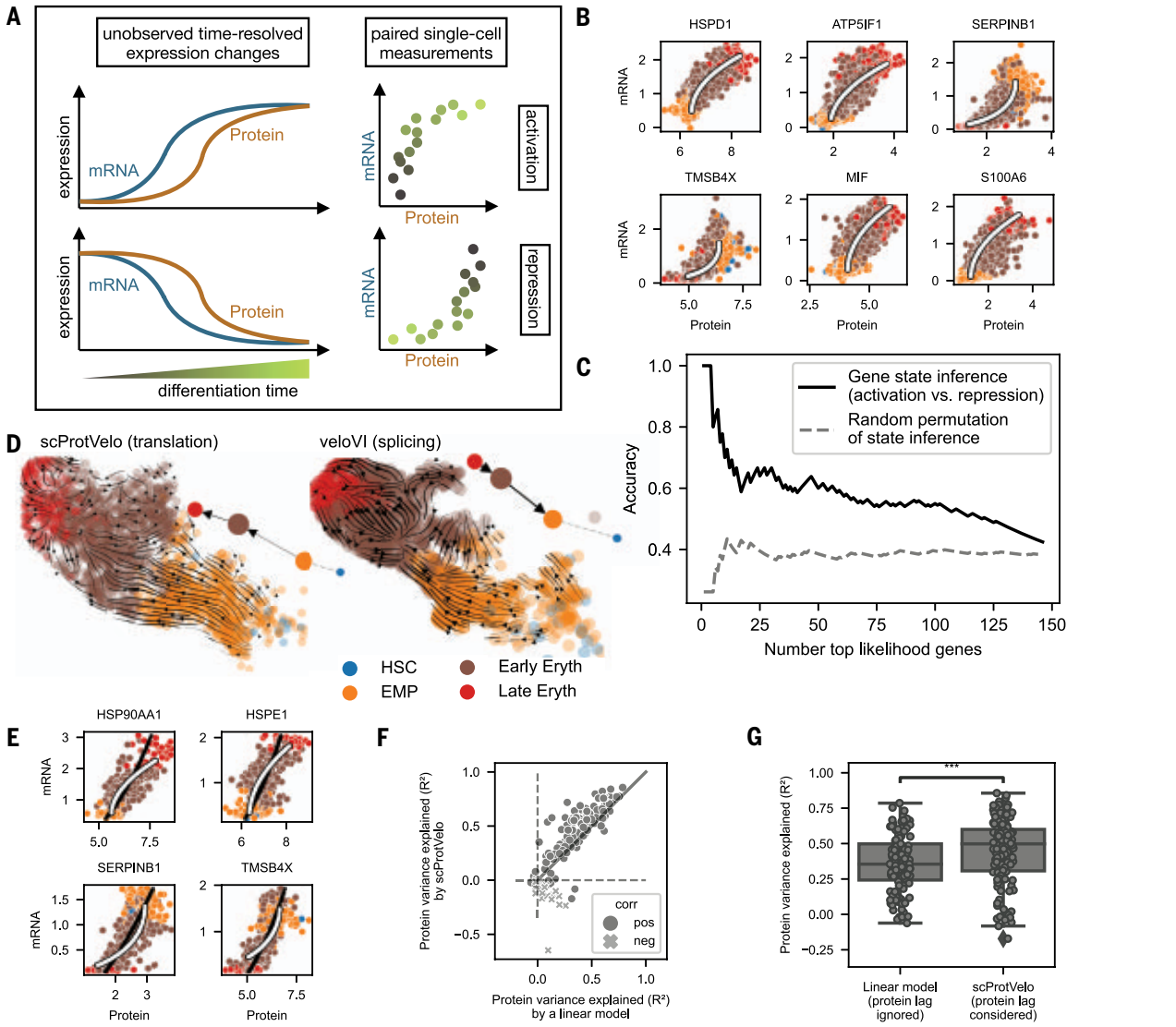


Fig. 7. scProtVelo models translation dynamics in cellular differentiation. (**A**) mRNA and protein abundance changes, time resolved versus observed in paired single-cell data. The time delay between the modalities allows the distinction of up- and downregulation in the data. (**B**) Exemplary model fits. (**C**) Performance of gene state predictions compared with uninformed predictions on genes that substantially change along the erythroid differentiation trajectory. (**D**) Translation-based (left) and splicing-based (right) velocities derived from the top 100 likelihood genes projected onto a UMAP embedding as well as the corresponding PAGA representations. (**E**) Exemplary comparison between fits from scProtVelo (white) and a linear model fitted in log space (black). (**F**) Gene-wise comparison of protein abundance variance explained by scProtVelo and a linear model, differentiated by the sign of the linear regression slope. (**G**) Global comparison of the protein abundance variance explained by scProtVelo and a linear model.

infer translation dynamics, we applied it to subsets of cells from the erythroid and pre-mDC differentiation trajectories, in which the assumption of constant kinetic rates per gene is reasonable, and compared the inferred gene states (activation versus repression) to ground truth defined by differential expression results between early and late cells. We indeed found the predicted dynamics to closely reflect the observed data (Fig. 7B and fig. S25, A and B) and model predictions of the gene state to be substantially better than random permutations in both differentiation branches (Fig. 7C and fig. S25C), implying that temporally delayed protein abundance changes were detectable. Further, aggregating the inferred protein velocity information of the top 100 genes recapitulated the correct cell progression in both trajectories (Fig. 7D, left, and fig. S25D). By contrast, applying the standard RNA velocity workflow to the scRNA-seq cells of the erythroid trajectory resulted in the previously reported erroneous backflow in velocity vectors from Late to Early Erythroid progenitors (*10*, *11*) (Fig. 7D, right). Also, judging individual proposed cell-to-cell transitions by whether they coincide with an increase in pseudotime consolidates higher accuracy in trajectory inference based on translation modeling than that on RNA velocity (fig. S25, E to I).

To quantify the impact of considering the temporal dynamics between mRNA and protein, we compared scProtVelo with a linear model, which cannot account for temporal delays. We considered 147 genes that were differentially expressed in early and late cells of the erythroid differentiation and incorporated pseudotime annotations as prior information to guide the inferred time into scProtVelo to ensure correct model fitting. Comparing the variance in protein abundance that can be explained through the corresponding mRNA expression, we found that scProtVelo resulted in more accurate modeling for almost all genes (except anticorrelated mRNA-protein pairs) (Fig. 7, E and F, and fig. S25B). Overall, accounting for translation dynamics when modeling protein abundance changes from mRNA levels led to a 40% increase in explained protein variance as compared with the simple assumption of a linear relationship (median R^2 values of 36 and 50% for the linear model and scProtVelo, respectively) (Fig. 7G). Taken together, our results demonstrate successful modeling of temporal gene expression dynamics from integrated scRNA-seq and scp-MS data, its potential for protein velocity-based trajectory inference, and its contribution to illuminating the often far from ideal mRNA-protein correlations.

Discussion

We presented an scp-MS dataset of the primary human HSPC compartment, encompassing >2500 cells that recapitulated the HSPC hierarchy through relative protein abundance changes. The combination with FACS allowed us to validate the scp-MS–defined cell states and revealed heterogeneity in surface marker–defined populations. Furthermore, integration of scp-MS with CITE-seq resulted in a joint latent space that enabled label transfer between modalities and a trajectory analysis that showed better performance compared with each of the two individual modalities. Vectors derived from the joint latent space, such as pseudotime or fate probabilities, enabled us to investigate differences between the modalities without the need for paired data. We found that the proteome level better described certain processes during HSC quiescence and differentiation, and we identified proteins which we experimentally validated to be important for the functional behavior of the most immature HSPCs.

In contrast to the HSC state, we found that the concordance between mRNAs and proteins is much higher during lineage specification, which may be partly explained by the higher effect sizes resulting in increased signal-to-noise measurements. Nonetheless, we were able to show that many of the proteins were not well explained by their mRNA profiles. Further, many of these proteins co-varied with functionally related proteins, highlighting the potential of scp-MS for revealing functional protein units and the importance of considering post-transcriptional regulation of protein abundances. For example, the abundance of B2M could be regulated on the protein level by protection from degradation through stabilization in the MHC I complex (*77*).

Finally, we investigated the temporal dynamics between the modalities by modeling translation dynamics. We found that this model could better explain the relationship between modalities than a linear regression, highlighting its relevance for revealing the often low correlations between the modalities. Further, we successfully used the learned dynamics to infer directionality of the pre-mDC and erythroid differentiation process, which for the latter has been challenging to achieve using splicing kinetics (*10*, *11*). Although the current necessary restrictions on the model complexity limit the simultaneous inference of multiple trajectories, we believe that this analysis serves as a compelling proof-of-concept and lays the groundwork for additional and more comprehensive studies of gene expression dynamics. Successfully uncovering the underlying processes of transcription, translation, mRNA, and protein degradation from single-cell data would open up a wide and highly impactful range of applications, especially in the context of in vivo cellular differentiation.

Although scp-MS is still associated with relatively low throughput and coverage, technological advances will continue to improve data depth and accuracy. In this context, our work offers a framework for more single-cell mRNA/protein-based multiomics studies, which have the potential to improve our understanding of cell states and translation dynamics across many biological systems including normal development, stem cell systems, and disease states such as cancer.

Materials and methods

Human bone marrow processing

Bone marrow samples were collected from healthy donors of European ancestry (young, 20 to 25 years; old, >60 years) following the standard protocol of the Department of Hematology, Rigshospitalet, with prior informed and written consent according to the Helsinki declaration under a protocol approved by the Danish National Ethics committee (approval number 1705391). Briefly, 15 mL of bone marrow was aspirated into a 20 mL syringe containing 5 mL of ice-cold sterile-filtered anticoagulant ACD, and immediately mixed to prevent clumping. The collected BM was then diluted 1:2 with PBS + 2mM EDTA + 0.5% BSA (PBS-EDTA) and filtered through a 70 μM filter. Subsequently, 30 mL of the diluted BM was gently layered over 15 mL of Lymphoprep (ProteoGenix, Cat. 1114544) and centrifuged at 400 g for 30 min at 20°C. The resulting supernatant, containing the interphase with the low-density mononuclear cells (LDMNCs), was transferred to a new tube. The LDMNCs were washed at a 1:1 ratio with PBS-EDTA, and the pellet was resuspended in 10 mL of PBS-EDTA. Cell counts were determined using a NucleoCounter, and the cells were frozen in FBS containing 10% DMSO overnight at −80°C before being transferred to liquid nitrogen storage. For thawing, cells were gently warmed in a water bath at 37°C for 2 to 5 min and then transferred to a 50 mL Falcon tube. 20 mL pre-warmed thawing medium (RPMI + 10% FCS + 100 μg/mL DNAse (Roche, Cat. 11284932001) + 5 mM $MgCl_2$) was added dropwise to the cell suspension, followed by centrifugation at 300 *g* for 10 min at room temperature. The supernatant was removed, and the cell pellet was resuspended in 10 mL of pre-warmed thawing medium and incubated for 15 min at 37°C. After the incubation, cells were counted and centrifuged.

FACS staining, analysis, and cell sorting

All FACS experiments included single-color controls using CompBeads (BD Biosciences, Cat. 552844) and Fluorescence Minus One (FMO) controls. Antibodies are shown in table S2.

scp-MS samples: Thawed BM mononuclear cells were CD34+ enriched using CD34 MicroBeads (Miltenyi Biotec, Cat. 130-046-702) according to the manufacturer's instructions. CD34+ HSPCs were subsequently incubated with a cocktail of biotinylated lineage antibodies and the

purified, unconjugated Endomucin antibody for 30 min on ice. After incubation, cells were washed with PBS-EDTA and centrifuged at 100 *g* for 10 min. Subsequently, cells were incubated for 30 min with Streptavidin-PECy5 and goat anti-rabbit IgG (PE), followed by another wash with PBS-EDTA at 100 *g* for 10 min. Finally, the cells were incubated with the panel of monoclonal antibodies for 30 min. Similar as previously described (*27*, *28*), after three washes with PBS to eliminate any remaining serum, cells were sorted on a BD FACSAria III instrument at a maximum speed of 500 events per second into Eppendorf twin.tec 384 LoBind plates containing 1 μL 20% TFE, 10 mM TCEP, 40 mM CAA, 50 mM TEAB lysis buffer. Live cells were sorted based on the membrane-impermeable viability dye 7-AAD. The 100 μm nozzle was used for all sorts into 384-well plates, and "Single Cell" and "4-way Purity" precision modes were selected for single cell and 500 cells/well carrier sorts, respectively. Immediately after sorting, plates were briefly centrifuged, snap-frozen on dry ice, and then boiled at 95°C in a thermocycler (Applied Biosystems Veriti 384-well) for 5 min. The plates were then briefly centrifuged again and stored at −80°C until further sample preparation.

CITE-seq samples: BM cells were incubated with the cocktail of biotinylated lineage antibodies and the panel of monoclonal antibodies for 30 min on ice. Live cells were sorted based on the membrane-impermeable viability dye 7-AAD. Cells were FACS-sorted as Lin–CD34+ into low-bind Eppendorf tubes using a BD FACSAria III instrument.

scp-MS sample preparation

Plates were thawed on ice and sonicated for 2 min in a water bath sonicator. Protein lysates from the single cells were digested with 2 ng of Trypsin (Promega, Cat. V5280), dissolved in 1 μl of 100 mM TEAB pH 8.5 containing Benzonase (Sigma, Cat. E1014) diluted to a ratio of 1:5000 (vol/vol) to digest any DNA that would interfere with downstream processing. For the carrier plate, containing 500 cells per well, the trypsin amount was increased to 10 ng in each well. All dispensing steps in this procedure were carried out using the Dispendix I-DOT One instrument. After dispensing the digestion buffer, the plates were centrifuged, vortexed, and incubated at 37°C overnight to complete the digestion.

On the next day, peptides were labeled with TMTPro 16plex reagents (ThermoFisher Scientific, Cat. A44520). 6 mM of each label (except 127C) in 1 μL acetonitrile (ACN) was added to the single cell wells, while the 500-cell carrier wells were labeled with 13 mM of TMTPro-126 reagent in each well. The plates were then incubated at room temperature for 1 hour. The labeling reaction was quenched by adding 1 μl of 1.25% hydroxylamine (ThermoFisher Scientific, Cat. 90115) in 100 mM TEAB to single cell wells and 2.5% to carrier wells, followed by 15 min incubation at room temperature. The carrier plate, containing peptides labeled with TMTPro-126, was pooled into one sample and desalted using a SOLAμ Solid Phase Extraction (SPE) column (ThermoFisher Scientific, Cat. 60209-001) following the standard cleanup protocol. Eluted and desalted peptides were concentrated to dryness in an Eppendorf Speedvac, after which they were resuspended in A* (2% ACN, 0.1% TFA). Subsequently, peptides from 14 single cells and 200-cell equivalent of the Lin-CD34+ carrier were pooled and desalted using a SOLAμ SPE 96-well plate, described in the following. Briefly, the SOLAμ plate was prepared by activating with 200 μL methanol, conditioning with 200 μL 80% ACN, 0.1% FA, and equilibrating with 200 μL 0.1% FA. Between each step, the plate was centrifuged at 700 *g* for 1 min. Next, 60 μL of 0.1% FA and 200-cell equivalent of the carrier (in 10 μL volume) were dispensed into each column. The final TMT sets were pooled from 14 single cells using a handheld multichannel pipette onto the prepared SOLAμ columns. After pooling, peptides were bound to the filter by centrifuging at 700 g for 1 min. The SOLAμ plate was washed twice with 200 μL 0.1% FA, and the peptides were eluted twice with 20 μL of 40% ACN, 0.1% FA into an Eppendorf Twin.tec 96-well LoBind plate. Finally, the peptides were dried at 60°C in an Eppendorf Speedvac and reconstituted in A* for individual mass spectrometry analysis.

CITE-seq sample preparation

Sorted Lin-CD34+ cells were washed with PBS and stained with Fc block for 10 min and then washed again. After Fc block the cells were stained for 30 min with the CITE-seq antibodies (table S3). After the staining, cells were washed once again, counted to reach the proper concentration, and loaded into the 10x instrument according to the manufacturer's protocol. Single-cell RNA sequencing was performed using the 10x Genomics 3′ v3 platform. Approximately 7000 to 10,000 cells were loaded in each 10x channel, after which the single-cell libraries were prepared according to manufacturer's instructions, except for the added CITE-seq steps. A specific primer was added in the cDNA amplification step and instead of discarding the small fragments in the bead purification, they were kept and used for the CITE-seq/ADT library preparation according to the original publication (*22*). After all libraries were ready, their quality and size were verified in the Bioanalyzer before sequencing on the NextSeq 500 Illumina platform. Two to three samples were pooled and sequenced together in each sequencing run.

CRISPR/Cas9 targeting in human CD34+ BM cells

For each donor, normal BM was aspirated and the CD34+ cells were isolated as described above. Cells were cultured for 2 days in StemSpan (Stem Cell Technologies, Cat. 09655) supplemented with TPO (100 ng/mL), SCF (100 ng/mL), FLT3 (100 ng/mL), IL6 (100 ng/mL), and UM171 (35 nM) and electroporated with the ribonucleoprotein complexes (RNPs) made with 0.3 μl recombinant Cas9 (10 μg/μl) using sgRNAs harboring 2'-O-Methyl at the first 3 bps and last bp, 3′phosphorothioate bonds between the first 3 bps and last 2 bps as well as an 80-mer SpCas9 scaffold (Synthego). For the control locus AAVS1 we used a single sgRNA (5′-GGGGCCACUAGGGACAGGAU-3′), whereas we used pairs of sgRNAs for targeting H1F0 (5′-GAATTCCACGTCCGCCCCTG-3′ and 5′-UUGGAGGCCUUGGCCCGCUU -3′), SOD1 (5′-UAAUGGACCAGUGAAGGUGU-3′ and 5′-CAUGAACAUGGAAUCCAUGC-3′), SOD2 (5′-GGUUGUUCACGUAGGCCGCG-3′ and 5′-CGUCACCGAGGAGAAGUACC-3′), and TALDO1 (5′-GCCGACACGGGCGACUUCCA-3′ and 5′-UCCAUCCUCUGACGCUUCAC-3′). Electroporation was carried out on an Amaxa D4 electroporator using the program DZ-100.

Edited regions were amplified from gDNA from targeted cells using the following primer pairs (left and right): AAVS1: 5′-AGGGAGTTTTCCACACGGAC-3′ and 5′-CAAAGTACCCCGTCTCCCTG-3′, SOD1: 5′-ACTCACTTTGCTTGCCACTC-3′ and 5′-TTGGGAGGTCGAGGAGGATA-3′, SOD2: 5′-ACTCACTTTGCTTGCCACTC-3′ and 5′-TTGGGAGGTCGAGGAGGATA-3′, TALDO1: 5′-GCTGAGGACCTGGGGAGA-3′ and 5′-ATGAAGGTGCACAGAGGGG-3′, H1F0: 5′-AGGAAAAGCAGCGACTCCT-3′ and 5′-TAGGAGACAGGAGGGAGTGT-3′. Sanger sequencing was carried out using the left (SOD1, SOD2 and TALDO1) or the right (AAVS1 and H1F0) primer, and Variant Allele Frequency (VAF) was assessed against a reference (WT) amplicon using the ICE analysis tool (ice.editco.bio).

Long-term culture initiating cell (LTC-IC) assay

One day after electroporation, CD34+ cells were sorted (10, 30, 100, 300, and 1000 cells; 10 wells for each number) onto Mitomycin-C treated MS5 feeder cells in 96-well plates and cultured for 5 weeks in Gartner's medium (αMEM supplemented with 11% FCS, 12.5% horse serum, 1% Pen/Strep, 1 μM hydrocortisone, 57.2 μM beta-mercaptoethanol, 20 ng/mL TPO, 20 ng/mL IL-3, and 20 ng/mL GCSF) followed by 2 weeks in methylcellulose (MethoCult, Stem Cell Technologies, Cat. H4435). Stem cell frequencies were assessed by counting wells positive for cell growth.

Co-culture assay

One day post-electroporation, CD34+ cells were sorted, and 20,000 cells were seeded in triplicate onto Mitomycin-C treated MS5 feeder cells in 6-well plates and cultured for 3 weeks in Gartner's medium (described in detail above).

FACS analysis of co-culture experiments

After 3 weeks co-culture (see above), adherent cells were trypsinized and stained with the following antibody conjugates: hCD45-BV421 (BD Horizon, Cat.563879, RRID:AB_2744402), CD41a-PECy7 (BD Pharmingen, Cat. 561424, RRID:AB_10642584), CD14-BV786 (BD OptiBuild, Cat. 741034, RRID:AB_2740651), CD11b-BV711 (BD OptiBuild, Cat. 742641, RRID:AB_2740934), CD15-PE (BD Pharmingen, Cat. 582371, RRID:AB_11154049), CD16-BV605 (BD Horizon, Cat. 563172, RRID:AB_2744297), CD66b-APC (BioLegend Cat. 305118, RRID:AB_2566607), and CD34-APC-CY7 (BioLegend, Cat #343514, RRID:AB_1877168). Finally, 7-AAD was added as a viability marker, and the cells were analyzed on a BD LSRFortessa X-20.

Colony forming cell (CFC) assay

One day post-electroporation, 1000 cells were plated in triplicate in 1 mL methylcellulose (MethoCult, Cat. H4435) and cultured for 2 weeks. Subsequently, different classes of colony-forming units (CFUs) were counted under a microscope.

Mass spectrometry data acquisition

Peptides were loaded onto a 2 cm C18 trap column (ThermoFisher 164946), connected in-line to a 15 cm C18 reverse-phase analytical column (ThermoFisher EasySpray ES904), with 100% Buffer A (0.1% Formic acid in water) at 4 µl/min or 750 bar maximum pressure using the ThermoFisher EasyLC 1200, and the column oven operating at 30°C. Peptides were eluted over a linear 140 min gradient at a flowrate of 100 nL/min, using 80% acetonitrile, 0.1% formic acid (Buffer B) going from 10% to 30% B. Spectra were acquired with an Orbitrap Eclipse Tribrid mass spectrometer with FAIMS Pro interface (ThermoFisher Scientific) running Tune 3.4 and Xcalibur 4.3 using the RETICLE acquisition method (*28*). FAIMS switched between compensation voltages of −50 V and −70 V with cycle times of 3 s and 1.5 s, respectively. MS1 spectra were acquired at 120,000 resolution with a scan range from 375 to 1500 m/z, normalized AGC target of 300%, and maximum injection time of 50 ms. Precursors were filtered using monoisotopic peak determination set to peptide, charge state 2 to 6, dynamic exclusion of 120 s with ±10 ppm tolerance excluding isotopes and different charge states, and a precursor fit of 70% in a window of 0.7 m/z. Precursors selected for MS2 analysis were isolated in the quadrupole with a 0.7 m/z window. Ions were collected for a maximum injection time set to 46 ms and normalized AGC target of 300%. Fragmentation was performed with 30 normalized CID collision energy, and MS2 spectra were acquired in the linear ion trap (LIT) at scan rate rapid. LIT MS2 spectra were subjected to real-time search (RTS) using the homo sapiens database obtained from Uniprot (Swiss-Prot with isoforms, 42,297 sequences, downloaded on 30.03.2021), and trypsin set as enzyme (cleavage next to arginine or lysine, but not before proline). Static modifications were TMTpro16plex on Lysine (K) and N-Terminus and carbamidomethyl on cysteine (C). Oxidation of methionine (M) was set as variable modification. Maximum missed cleavages were set to 1 and maximum variable mods to 2. FDR filtering was enabled, maximum search time was set to 40 ms, and the scoring threshold was set to 1.4 XCorr, 0.1 dCn, and 10 ppm precursor tolerance. Use as trigger only was checked, and closeout was enabled with maximum number of peptides per protein set to 4. Precursors identified via RTS were isolated in the quadrupole with a 0.7 m/z window, ions were collected for a maximum injection time of 750 ms and normalized AGC target of 500%, fragmented with 32 normalized HCD collision energy, and MS2 spectra were acquired in the orbitrap at 120,000 resolution with first mass set to 120 m/z.

Mass spectrometry data processing

Raw files were analyzed with Proteome Discoverer 3.1 (ThermoFisher Scientific) with the built-in TMTpro Reporter ion quantification workflows using the standard settings if not further specified. Reporter ion quantification was performed only on FTMS MS2 spectra, and the same spectra were also sent to Sequest for identification, where they were searched with precursor mass tolerance of 10 ppm and fragment mass tolerance of 0.02 Da. The database was the homo sapiens database obtained from Uniprot (Swiss-Prot with isoforms, 42,412 sequences, downloaded on 10.03.2024) with trypsin set as enzyme (cleavage next to arginine or lysine, but not before proline) and maximum number of missed cleavages set to 2. Static modifications were TMTpro16plex on lysine (K) and N terminus, and carbamidomethyl on cysteine (C) was set. Dynamic modifications were set as oxidation (M) and Met-loss on protein N-termini. Results were rescored with Inferys (*78*) and q-values were assigned with percolator and filtered to 1% FDR. For the reporter ion quantification, normalization mode and scaling mode were set to none and average reporter s/n threshold was set to 0. Isotopic error correction was applied.

CITE-seq data processing

Demultiplexed FASTQ files were aligned to the human reference genome (GRCh38, v 2020-A as provided by 10x Genomics, which is based on GENCODE v 32/Ensembl 98) using Cell Ranger software (v 5.0.1) by 10x Genomics. The Cell Ranger "count" standard pipeline was used to obtain the expression matrix of unique molecular identifiers (UMIs) for each individual sample.

scp-MS data analysis

scp-MS data were analyzed using the SCeptre pipeline (*27*) using Python 3.9 with SCeptre version 1.1.0, flowkit 0.9.3 (*79*) and Scanpy 1.9.1 (*29*). FACS .fcs files were processed in flowkit in the following manner. First, compensation and bi-exponential transform was applied and subsequently, data was merged and batch corrected using combat (*80*) with each individual representing a batch. Subsequently, each FACS parameter in the merged data was transformed across single cells using the robust z-score with value clipping from −5 to 5. The resulting data were used to set the gates for the populations to annotate cells with a gated population. FACS data and sort- and label layouts were used to create the metadata for each cell.

Subsequent SCeptre analysis was performed in the following manner. A first pass of the SCeptre workflow was applied to identify clusters of cells that likely contain low quality cells (fig. S1A). Indications for low quality cell clusters were mixing of cells of unrelated FACS populations, low total protein signal and the combination of low FSC-A with high SSC-A. Cells contained in the low-quality clusters were filtered out in the second pass of the SCeptre workflow, which included the following steps. Raw files with less than 2000 peptide-spectrum matches (PSMs) were removed and cells from low quality clusters were removed. To remove proteins whose values likely stem from contamination, a mean single cell to carrier ratio was calculated. For each protein in each cell, the signal ratio to the corresponding carrier channel was calculated. Proteins with a mean ratio of over 0.2 were removed. This also removed proteins where the ratio could not be calculated since there was no single cell signal in any cell. Subsequently, protein abundance data were batch corrected using the SCeptre normalize function. Cells were filtered to have a signal for at least 600 proteins, and a minimum log2 total signal of 13.6. Furthermore, cells were filtered based on a relationship of FSC-A to log2 total signal (Fig. 1E), as these parameters showed high correlation and thus were deemed to be suitable to filter outlier cells. The relationship between the two parameters was modeled using the function sklearn.linear_model.RANSACRegressor and cells with a prediction error of over 0.8 log2 total signal were removed. Finally, proteins were required to have a signal in at least 10 cells. This resulted in a dataset of 2934 proteins across 2506 cells with various data completeness (Fig. 1C) and for the second dataset 2174 proteins across 922 cells.

Subsequently, protein abundance data were median-ratio normalized and log2(x+1) transformed. This represents the "batchcorr_norm_log2" layer that was used for differential expression testing. For the

embedding, data were imputed using the k-nearest neighbor algorithm followed by the application of the "regress_out" function of Scanpy using log2 total signal and number of proteins with signal as parameters. To embed the data, principal component analysis (PCA) was applied and used to calculate a neighborhood graph, which was subsequently embedded using UMAP. Unsupervised cell clustering was performed using the Leiden algorithm. Clusters were manually annotated using the abundance of marker genes and FACS information.

The quality of batch effect removal was quantified as the overall variance contribution given a covariate (PC regression value) as implemented in the python single-cell integration benchmarking package scib (*32*). Since the SCeptre batch correction is based on alignments of medians and thus assumes balanced multiplexed sets, we expected good integration of the actively balanced multiplexed sets "Enriched Populations". The observation that the batch correction also worked for the "Total HSPC" sets indicated that random samples of 14 HSPCs were sufficiently balanced.

Differential expressed proteins in each cluster were calculated by comparing protein abundance in each cluster against all other clusters using the layer "batchcorr_norm_log2". For each protein, a two-sided Welch's *t* test was performed, excluding missing values. The test was chosen since we do not assume equal variance and sample size between groups. P-values were corrected via the Benjamini–Hochberg procedure and a cutoff of 5% FDR was applied.

To compare protein log2 fold changes derived from scp-MS to bulk data, we used proteomics data (*42*) from bulk sorted human bone marrow populations, which were reanalyzed with Spectronaut 18.4 using factory settings. Replicates were aggregated to the mean and the log2 fold change of each protein in each population was calculated by subtracting the mean of all other populations. scp-MS data were matched to the smallest matching bulk protein group. For each scp-MS cluster and gated population, log2 fold changes were compared for proteins that had at least 30% coverage in the scp-MS cluster and were detected in the bulk dataset. Each scp-MS cluster and gated population was compared to each bulk population by calculating the Pearson correlation and the resulting matrix was clustered hierarchically.

For unsupervised clustering of proteins, proteins were filtered to have at least one significant log2 fold change above 0.2 (30% cluster coverage, 1% FDR) in any of the cell clusters. For each cell cluster, values from layer "batchcorr_norm_log2" were aggregated to the mean ignoring missing values and each protein was z-score transformed across cell clusters. Proteins were clustered hierarchically using Euclidean distance as metric and the linkage method ward and the tree was cut to 9 protein clusters. Each protein cluster was tested for enriched GO Biological Process and KEGG terms using gseapy (*50*). The background was set to all proteins in the dataset and terms were filtered 5% FDR and more than one term per gene set. For each protein cluster selected terms were plotted.

CITE-seq data analysis

The CITE-seq expression matrices were analyzed using scanpy 1.9.8 (*29*) and scvi-tools 1.1.2 (*53*). Cells were filtered to contain at least 1800 mRNA counts, at least 800 genes, and a fraction of mitochondrial gene counts between 0.025 and 0.1. Genes were filtered to be detected in at least 99 cells. Cells were annotated via Azimuth (*4*), separately for each individual. We removed cells that were annotated as a rare celltype or celltypes with no evidence in the scp-MS data. Only CITE-seq surface markers that overlapped with scp-MS markers were retained. Genes were mapped to the scp-MS data using an ENSG-Uniprot mapping table. If multiple proteins or mRNAs mapped to the same ENSG, the protein or mRNA with the highest coverage was mapped. Subsequently, we created the "counts_totalnorm_log2" layer by normalizing each cell to total counts and applying log(x+1) transform. Next, genes were subsetted to contain 4000 highly variable genes with flavor "seurat_v3" and batch keys provided, and to contain all genes detected in scp-MS, resulting in 5820 genes. Count data were modeled and batch-corrected using totalVI (*52*). The modeled mRNA expression was median-ratio normalized and log2 transformed, resulting in the "denoised_rna_norm_log2" layer. For the UMAP embedding, the modeled mRNA expression was used. Azimuth GMPs were subsetted into GMDP and MDP using leiden clustering. Cell cycle phase was predicted using a list of marker genes and the "score_genes_cell_cycle" function of Scanpy. To benchmark mRNA log2 fold changes from either counts or scVI-modeled data, transcriptomics data (*42*) from bulk sorted human bone marrow populations were used. Transcriptomics data were CPM normalized and log(x+1) transformed. Log2 fold change between two populations was calculated by subtracting the means of the replicates.

Integration of scRNA-seq and scp-MS

Integration was performed using GLUE (*56*). We linked mRNA features from the CITE-seq data to their corresponding proteins in the scp-MS data. The guidance graph was constructed to contain all 2934 proteins of scp-MS and all 5820 mRNAs from scRNA-seq (CITE-seq), resulting in 8754 nodes. 2778 edges with weight 0.9 were inserted between protein and matching mRNA. Self-loop edges with weight 1.0 were inserted for each node. This resulted in 11,532 edges. Further required input for GLUE were an expression matrix and a latent representation thereof per modality. For scp-MS, the imputed data that were not processed with "regress_out" were used as expression matrix and as latent representation, 50 PCs of that were used. The probabilistic generative model used by the decoder was set to "Normal" corresponding to a "LogNormal" distribution considering the prior log-transformation of the scp-MS data. For scRNA-seq, the count data were used as expression matrix and the totalVI latent space was used as representation, which provided additional batch correction. The probabilistic generative model was set to negative binominal ("NB") and the individuals were defined as batch variable. In contrast to the standard GLUE workflow that used a two-stage training procedure to estimate balancing weights, we used a single-stage procedure without balancing weights. The GLUE integration resulted in a joint latent representation of "protein" and "mRNA" cells. After integration, scRNA-seq labels were transferred to scp-MS using the GLUE function "transfer_labels" using the GLUE latent space and 15 nearest neighbors. Modality integration as well as biological signal preservation were measured using scIB metrics (*32*). To quantify the overlap of the distribution of cells from the two modalities in the integrated latent space, we used the average silhouette width score that compares the average distance between cells of the same modality to the average distance of cells from different modalities. The highest possible score for separation of the modalities is 1, while values near 0 indicate well overlapping sets of cells. To assess preservation of biological variance, we compared cell type separation metrics before and after integration per modality. For the "mRNA" cells, we used the Azimuth-derived cell type annotations as ground truth that are to be preserved and the totalVI embedding as baseline. For the "protein" cells, we used the Leiden cluster-based cell types and the PCA embedding that was also used as input to GLUE. We computed the average silhouette width score as just described as well as the normalized mutual information and adjusted rand index scores as implemented in scIB. Both compare the overlap of two clusterings, one being the ground truth cell type annotations and the other one being derived from the embedding structure as data-driven clusterings. A perfect match of the clusterings would lead to a score of 1, uncorrelated or random clusterings to a score of 0.

Cellrank analysis

A cellrank (*5*, *6*) analysis was performed on the integrated dataset, with standard setting applied if not further specified. First, the joint latent space was used to calculate a nearest neighbor graph, following a diffusion map. Then, after setting an immature HSC as root cell, a diffusion pseudotime was calculated, which was used to initialize a cellrank

pseudotime kernel. Subsequently, a transition matrix was computed and macrostates were determined via GPCCA. Terminal states were determined by fitting seven macrostates to a Schur decomposition and selecting the top six. Finally, fate probabilities were calculated for the six terminal states.

Expression profiles were modeled using the approach in cellrank where a generalized additive model (GAM) is fitted to the expression values along pseudotime, while using the fate probabilities as weights. Cells were subsetted into the relevant clusters for each trajectory. GAM settings were 12 kn, lam=12 and derivative was set as penalty. For mRNA, the scVI-modeled values were used. For protein, missing values were ignored by setting the weights of cells with missing values for the specific protein to zero. For both modalities, values were not log2 transformed, but used in linear space. Furthermore, values were normalized by dividing them by the global mean of the specific gene.

HSC quiescence and differentiation was investigated using a subset of cells with pseudotime below 0.4. Each protein or mRNA had to be detected in at least 200 cells to be considered. Next, the correlation vector was calculated by calculating the Pearson correlation of each log2 transformed mRNA/protein with pseudotime. For protein, missing values were ignored for this calculation, while for the count-based mRNA, zeros were used. scVI modeled mRNA did not have any missing values.

Lineage specification was investigated using all cells, however using only mRNA/proteins with detections in at least 200 cells. Correlations vectors for each lineage were calculated by calculating the Pearson correlation of each log2 transformed mRNA/protein with fate probability. For protein, missing values were ignored, while for the count-based mRNA, zeros were used. scVI modeled mRNA did not have any missing values.

Gene set enrichment analysis (GSEA) was performed on the correlation vectors using GSEApy (*50*). A preranked GSEA was computed using the gene set GO Biological Process with minimum set size of 3 and maximum of 1000. Resulting terms were filtered to 10% FDR and for the lineage specification additionally to a positive enrichment score. The term list was shortened by removing redundant GO term using REVIGO (*81*).

Gene expression cascades for each trajectory were visualized as heatmaps by filtering the correlation vector to a minimum correlation of 0.25 for mRNA and 0.2 for protein. Then genes were sorted by the pseudotime coordinate of their maximum expression in the modeled abundance profile. Then, this vector was binned into 4 bins and the top 5 mRNAs or top 8 proteins of each bin were selected for visualization.

For the joint mRNA-protein covariation analysis the protein lineage specification correlation vectors were concatenated into a table of protein lineage correlations across the six lineages. In each vector column, if the protein did not pass the 200-cell coverage filter, it was set to missing. The table was subsequently filtered for proteins with Pearson correlation of at least 0.2 in any of the six lineages. For mRNA the count-based correlation vectors were used. Then the intersection of both protein and mRNA tables was concatenated and a pairwise correlation matrix was computed, while requiring at least four values to return a Pearson value. Keeping only columns with proteins yielded a correlation vector of a protein with all other mRNAs and proteins and sorting this vector by correlation resulted in a correlation rank for each mRNA and protein. For each protein, related proteins were defined to be in the same KEGG pathway.

To compare the mRNA correlation ranks for each protein to results from bulk data, we used a dataset containing mRNA and protein measurements of 59 breast cancer cell lines (*74*–*76*). Log2 transformed tables were normalized by subtracting the row and column medians. Similar to the single-cell data, tables were concatenated, and a correlation matrix was computed and an mRNA correlation rank for each protein was computed. Ranks were then compared between bulk and single-cell data.

Gene expression modeling with scProtVelo

Creating paired single cell data: To create a dataset with paired mRNA and protein expression for all cells, we selected cells of the erythroid and pre-mDC trajectories based on a leiden clustering of the GLUE embedding, so that we obtain the same biological subset of cells from both modalities. For both trajectories, we then removed cells with annotated cell types for which we don't know ground truth transitions. The resulting cell subsets contained 472 cells from the protein dataset and 1677 cells from the scRNA-seq data for the erythroid branch and 262 cells from the protein dataset and 897 cells from the scRNA-seq dataset for the pre-mDC branch. We selected proteins that were detectable in at least 30% of the protein cells within the subsets and mRNAs that had nonzero counts in at least 2% of the scRNA-seq cells. This resulted in 1065 proteins and 5101 mRNAs for the erythroid subset, which gave us a total of 1036 overlapping genes to work with. For the pre-mDC subset, we got 1101 proteins and 5081 mRNAs with an overlap of 1072 genes. mRNA counts and batch corrected proteins were median-ratio normalized and, as also done for RNA velocity, expression values were smoothed per modality over their respective nearest neighborhood graphs. Then, for each cell, values for the not measured modality were interpolated over the 15 nearest neighbor cells of the modality of interest based on the joint GLUE embedding. For all features, 25% and 75% quantiles were computed and used to scale features by $1 / (75\%\ quantile - 25\%\ quantile)$ to give all features the same weight in the scProtVelo reconstruction loss later on. We want to note here that scaling of expression values only results in scaled transcription and translation rates that can be reverted after model fitting, while a shift cannot be compensated for within the differential equations we define below.

We further define genes to showcase substantial variation during the selected differentiation process, if they are differentially expressed between the upper and lower 30% quantiles of cell with respect to diffusion pseudotime in both modalities. For that we randomly sampled 50 cells from the respective quantiles, performed an unpaired t test and selected genes at a significance level of 5% after Benjamini-Hochberg multiple testing correction that had a log-fold change of at least 0.1 on protein level and 0.5 on mRNA level. This resulted in a selection of 146 genes for the erythroid subset and 46 genes for the pre-mDC branch.

The translation dynamics model: In a first modeling setup, we followed the concept of RNA velocity (*7*). We jointly model mRNA $r(t)$ and protein $p(t)$ expression of a gene considering the transcription of a gene to mRNA molecules at rate α, their degradation at rate β as well as translation to proteins at rate κ and their respective degradation at rate δ over time t. We assume fixed (time independent) rates per gene and also model time as gene specific, since orthogonal processes like cell cycle as well as cell differentiation could be the driving factors in different genes. Per gene, this process can be formalized by the following ordinary differential equations:

$$\frac{dr(t)}{dt} = \alpha - \beta r(t)$$

$$\frac{dp(t)}{dt} = \kappa r(t) - \delta p(t)$$

The expression values $r(t)$ and $p(t)$ stand for nearest-neighbor smoothed median-ratio normalized abundances. Both, the time annotations t as well as the kinetic rates α, β, κ and δ are not directly measured and are to be inferred.

Given that we model fixed, time-independent rates, a gene can either be in an induction or repression state, characterized by mRNA expression increasing or decreasing with time. In the RNA velocity formulation, induction is defined to start with unspliced as well as spliced

expression being 0. We observe different behavior in mRNA-protein phase portraits. Thus, we assume free initial expression values ${r_0}^{ind}$ and ${p_0}^{ind}$ and pose the restriction on the kinetic parameters that the derivatives $\frac{dr(t)}{dt}$ and $\frac{dp(t)}{dt}$ are nonnegative at ${r_0}^{ind}$ and ${p_0}^{ind}$, respectively. For the case of repression, we model the initial expression values ${r_0}^{rep}$ and ${p_0}^{rep}$ as a result of an unobserved induction phase that shares all kinetic parameters, but the transcription rate, with the following repression phase and enforce that the derivative of the mRNA expression is non-positive at the inferred starting point $({r_0}^{rep},\ {p_0}^{rep})$. Additionally, similar to the RNA velocity model, we explicitly also allow cells to be in a steady state that we set to be the respective r_0 and p_0 values depending on the general gene state.

On top of this very free model setup, we used a second model variation that is meant to capture one known linear process more accurately. In this setting, we restrict time to be shared across all genes and introduce diffusion pseudotime as a prior on the modeled time.

scProtVelo: We implemented all those considerations into scProtVelo, which is available as a Python package via github.com/theislab/scProtVelo. The scProtVelo model, in short, is a variational autoencoder-like architecture that takes a cell representation as input, learns a latent representation of it that is further used to generate mRNA and protein expression that is close to the original expression of that cell. The crucial part of scProtVelo is that the generation is not performed by a fully connected neural network. Instead, it infers a latent time annotation for a cell as well as the kinetic parameters and combines both to generate the expression values based on the defined expression dynamics. Our python implementation heavily builds on the veloVI framework, and we refer the reader to the original publication (*9*) for all technical modeling details that are not explained in the following.

The scProtVelo generative model: The differential equations introduced above can be solved analytically and as

$$r(t) = r_0 e^{-\beta t} + \frac{\alpha}{\beta}\left(1 - e^{-\beta t}\right)$$

$$p(t) = p_0 e^{-\delta t} + \frac{\alpha\kappa}{\beta\delta}\left(1 - e^{-\delta t}\right) + \frac{\kappa\left(r_0\beta - \alpha\right)}{\beta(\delta - \beta)}\left(e^{-\beta t} - e^{-\delta t}\right)$$

with initial expression values r_0 and p_0 and as such build the base for the generative process of scProtVelo. We can now formalise all four possible cell-gene states (induction steady state, induction dynamic state, repression steady state, repression dynamic state) described by the gene specific variable $k^{ind/rep}$ and the gene and cell specific variable ${k_i}^{ss/ds}$ for a cell $i \in I$. Time t as well as kinetic rates $\alpha,\ \beta,\ \kappa$ and δ and intermediate variables are gene specific, but we omit the subscripts for readability. For the induction steady state, we propose a restriction-free expression pair

$$r\left(\cdot, k^{ind/rep} = 0, {k_i}^{ss/ds} = 0\right) := {r_0}^{ind}$$

$$p\left(\cdot,\ k^{ind/rep} = 0,\ {k_i}^{ss/ds} = 0\right) := {p_0}^{ind}.$$

For the corresponding dynamic case, mRNA and protein expression can be described to follow the analytical solution given above with the initial state being the values defined for the steady state case:

$$r\left(t_i, k^{ind/rep} = 0, {k_i}^{ss/ds} = 1\right) := {r_0}^{ind} e^{-\beta t_i} + \frac{\alpha}{\beta}\left(1 - e^{-\beta t_i}\right)$$

$$p\left(t_i,\ k^{ind/rep} = 0,\ {k_i}^{ss/ds} = 1\right) := {p_0}^{ind} e^{-\delta t_i} + \frac{\alpha\kappa}{\beta\delta}\left(1 - e^{-\delta t_i}\right) + \frac{\kappa\left({r_0}^{ind}\beta - \alpha\right)}{\beta(\delta - \beta)}\left(e^{-\beta t_i} - e^{-\delta t_i}\right).$$

For the repression modeling, we assume an unobserved induction phase preceding the observed repression phase. We assume that the induction phase shares rates $\beta,\ \kappa$ and δ with the repression phase and that both phases just differ in their transcription rates α_{ind} and α_{rep}. Further assuming an initial state 0, 0 for the induction and a gene-specific time $t_s \in \mathbb{R}_+$ at which the initial expression pair is attained, we propose the repression steady state to be

$$r\left(\cdot, k^{ind/rep} = 1, {k_i}^{ss/ds} = 0\right) := \frac{\alpha_{ind}}{\beta}\left(1 - e^{-\beta t_s}\right)$$

$$p\left(\cdot,\ k^{ind/rep} = 1,\ {k_i}^{ss/ds} = 0\right) := \frac{\alpha_{ind}\kappa}{\beta\delta}\left(1 - e^{-\delta t_s}\right) - \frac{\kappa\alpha_{ind}}{\beta(\delta - \beta)}\left(e^{-\beta t_s} - e^{-\delta t_s}\right).$$

Last, for the dynamic repression state, we obtain

$$r\left(t_i, k^{ind/rep} = 1, {k_i}^{ss/ds} = 1\right) := {r_0}^{rep} e^{-\beta t_i} + \frac{\alpha_{rep}}{\beta}\left(1 - e^{-\beta t_i}\right)$$

$$p\left(t_i,\ k^{ind/rep} = 1,\ {k_i}^{ss/ds} = 1\right) := {p_0}^{rep} e^{-\delta t_i} + \frac{\alpha_{rep}\kappa}{\beta\delta}\left(1 - e^{-\delta t_i}\right) + \frac{\kappa\left({r_0}^{ind}\beta - \alpha_{rep}\right)}{\beta(\delta - \beta)}\left(e^{-\beta t_i} - e^{-\delta t_i}\right)$$

with ${r_0}^{rep}$ and ${p_0}^{rep}$ being the values obtained above for the repression steady state.

For the generative process of an expression pair $\left(r_i, p_i\right)$ of a cell i, we draw a low-dimensional $(d = 10)$ latent cell representation

$$z_i \sim Normal\left(0,\ I_d\right)$$

as well as a cell-independent distribution over gene states (induction or repression) described by the probability simplex generated by the Dirichlet distribution and then the actual state assignment itself

$$\pi^{ind/rep} \sim Dirichlet(0.25,\ 0.25)$$

$$k^{ind/rep} \sim Categorical\left(\pi^{ind/rep}\right).$$

Further, we draw a distribution distinguishing between a cell being in steady state or dynamic state as well as the assignment

$${\pi_i}^{ss/ds} \sim Dirichlet(0.25,\ 0.25)$$

$${k_i}^{ss/ds} \sim Categorical\left({\pi_i}^{ss/ds}\right).$$

The time annotation t_i, we model as a function of the latent cell representation z_i

$$t_i = h\left(z_i\right)$$

with $h\colon\ \mathbb{R}^d \to (0,\ 1)$being a fully connected neural network.

Finally, the expression pair $\left(r_i, p_i\right)$ is drawn from two normal distributions as

$$r_i \sim Normal\left(r\left(t_i,\ k^{ind/rep},\ {k_i}^{ss/ds}\right),\ \sigma^2\right)$$

$$p_i \sim Normal\left(p\left(t_i,\ k^{ind/rep},\ {k_i}^{ss/ds}\right),\ \sigma^2\right).$$

scProtVelo inference: The goal of the inference is to obtain estimates for all introduced kinetic rates, gene state distribution parameters, neural network weights, initial expression values for the induction cases, switch times for the repression cases and finally posterior distributions over the latent variables $z,\ \pi^{ind/rep}$ and $\pi^{ss/ds}$. We use variational inference and with $x_{i,glue}$ being a precomputed representation of the mRNA and protein expression of cell $i \in I$, namely its GLUE representation, we propose the following factorization of the posterior distribution

$$q_\phi(z, \pi^{ind/rep}, \pi^{ss/ds}|x_{glue}) := \prod_i^I q_{1,\phi}(z_i|x_{i,glue}) \prod_g^G q_{2,\phi}(\pi_g{}^{ind/rep}, \pi_{ig}{}^{ss/ds}|z_i)$$

meaning that a neural network q_1 maps a precomputed cell representation $x_{i,glue}$ to a posterior distribution for the latent cell representation z_i which is itself used as input to a second neural network q_2 that maps it to posterior distributions for the state distributions $\pi_g{}^{ind/rep}$ and $\pi_{ig}{}^{ss/ds}$ for all genes $g \in G$. For the likelihoods we integrate over the four defined cell-gene states and obtain:

$$p_\theta(r_{ig}|z_i, \pi^{ind/rep}, \pi_i{}^{ss/ds}) = \sum_{k_g{}^{ind/rep}\in 0,1} \sum_{k_{ig}{}^{ss/ds}\in 0,1} \pi_g{}^{ind/rep} \pi_{ig}{}^{ss/ds} Normal(r(t_{ig}, k_g{}^{ind/rep}, k_{ig}{}^{ss/ds}), \sigma^2)$$

$$p_\theta(p_{ig}|z_i, \pi^{ind/rep}, \pi_i{}^{ss/ds}) = \sum_{k_g{}^{ind/rep}\in \{0,1\}} \sum_{k_{ig}{}^{ss/ds}\in \{0,1\}} \pi_g{}^{ind/rep} \pi_{ig}{}^{ss/ds} Normal(p(t_{ig}, k_g{}^{ind/rep}, k_{ig}{}^{ss/ds}), \sigma^2).$$

Additionally, to minimizing the arising ELBO objective that is composed of the negative log likelihood of the data given the fitted model as well as Kullback-Leibler divergence terms for the latent variables z, $\pi^{ind/rep}$ and $\pi^{ss/ds}$, we include two more loss terms to be optimized. First, we add $L_{param,ind}$ to enforce nonnegative derivatives in mRNA and protein expression at the initial values in the induction case and $L_{param,rep}$ to ensure non-positive derivatives in mRNA expression at the initial values in the repression case

$$L_{param,ind}(\theta, \phi; \{x_{i,glue}, r_i, p_i, i\in I\}) := \frac{1}{|G|}\left(\sum_g^G ReLU\left(r_{g,0}{}^{ind} - \frac{\alpha}{\beta}\right) + \sum_g^G ReLU\left(p_{g,0}{}^{ind} - \frac{\kappa}{\delta} r_{g,0}{}^{ind}\right)\right)$$

$$L_{param,rep}(\theta, \phi; \{x_{i,glue}, r_i, p_i, i\in I\}) := \frac{1}{|G|} \sum_g^G ReLU(\alpha_{rep} - \beta r_{g,0}{}^{rep})$$

with loss values set to 0 whenever a denominator is 0 and weighted strongly enough to enforce all wanted conditions.

Secondly, we introduce $L_{init,ind}$ and $L_{init,rep}$. as penalty terms used during a pretraining phase that encourages the modeled steady states to coincide with heuristically observed steady state values $r_g{}^{high}$, $r_g{}^{low}$, $p_g{}^{high}$ and $p_g{}^{low}$ that are computed as the observed median expression of the cells that have mRNA expreion within the highest or lowest 1% quantile of that gene

$$L_{init,ind}(\theta, \phi; \{x_{i,glue}, r_i, p_i, i\in I\}) := \frac{1}{|G|} \sum_g^G \left(r_{g,0}{}^{ind} - r_g{}^{low}\right)^2 + \left(p_{g,0}{}^{ind} - p_g{}^{low}\right)^2$$

$$L_{init,rep}(\theta, \phi; \{x_{i,glue}, r_i, p_i, i\in I\}) := \frac{1}{|G|} \sum_g^G \left(r_{g,0}{}^{rep} - r_g{}^{high}\right)^2 + \left(p_{g,0}{}^{rep} - p_g{}^{high}\right)^2.$$

Trajectory inference using scProtVelo: In the trajectory inference setting, we start model fitting with a pretraining phase with a high focus on the L_{init} loss terms, because we noticed that it helped general model fitting to get initial states approximately correct first. During normal training, this loss term is then turned off.

To quantify the performance in inferring the correct direction of cell progression, we subsetted all genes we fitted the model on to the ones that we accepted as changing along the differentiation trajectory and computed the direction of change as the sign of the difference between mean expression within the late 20% quantile with respect to pseudotime and the mean within the early 20% quantile of the cells. For the predicted state, we chose the state that had the higher value in $\pi_g{}^{ind/rep}$. To get an impression of the model performance, we sorted the genes by their model likelihood which represents a measure of how well the model captured the expression values of a gene, without knowing whether the direction of change was inferred correctly or not. Finally, we computed accuracy scores on increasing numbers of genes and compared it to the expected accuracy of a random permutation of the model predictions.

We based velocity computations on the 100 genes with the highest likelihood values and followed the same procedure as done for RNA velocity in the veloVI framework.

For the comparison to RNA velocity results, we followed the proposed veloVI workflow on unspliced and spliced mRNA counts of the mRNA cells that were used in the scProtVelo run. To create the .loom files, velocyto 0.17.17 (7) was applied on the cellranger count output. To compare the accuracy of the cell-to-cell transition graphs that the stream plot visualizations are based on, we checked whether proposed transitions from one cell to a neighboring cell coincide with an increase in diffusion pseudotime or not. We aggregated these values to accuracies per cell that we inspected visually projected on the respective UMAP embeddings as well as compared their distributions in a histogram.

Gene expression modeling with scProtVelo: To quantify the variance in protein expression that is determined by the corresponding mRNA expression, we figured it made only sense to consider genes that exhibit biological variation and restricted our analysis on the genes we found to be differentially expressed between the start and the end of the chosen differentiation trajectory. To ensure further that we quantify the variance corresponding to cell differentiation, we restricted latent time inference to only one shared time shared across all genes guided by a diffusion pseudotime prior.

To accommodate for those model specification changes, we first included a loss term representing the diffusion pseudotime prior $t_i{}^{dpt}$, $i \in I$ as

$$L_{time}(\theta, \phi; \{x_{i,glue}, r_i, p_i, t_i{}^{dpt}, i\in I\}) := \sum_i^I \left(t_i - t_i{}^{dpt}\right)^2.$$

Also, the shared time stands in conflict with the prior implementation of the steady states. So instead of modeling a steady state vs dynamic state variable per cell and gene, we learn a switch time per gene at which gene expression for a gene switches from a constant steady state value to dynamic expression. For induction state, this switch time is restricted to be within $(0, 1)$ just as the fitted time. For the repression case, we model a switch time in $(-\infty, 1]$, where a negative value implies an unobserved earlier initial state and observed cells just cover a more advanced time range. This is essential considering that the initial state for repression is tied to an induction phase. And finally, having access to diffusion pseudotime, we also refine on the heuristically estimated initial states and use the median expression of the cells within the 1% quantile with respect to diffusion pseudotime.

During training we leave out 20% of the cells at random as a test set. To compare protein estimates from scProtVelo against a linear model, we select the protein cells to avoid the computationally interpolated protein values. We then log transform mRNA and protein expression values and fit a linear model (including a bias term) on the train set. Finally, we compute R^2 values for protein predictions from both models in log expression space.

REFERENCES AND NOTES

1. E. Laurenti, B. Göttgens, From haematopoietic stem cells to complex differentiation landscapes. *Nature* **553**, 418–426 (2018). doi: 10.1038/nature25022; pmid: 29364285
2. J. D. Buenrostro *et al.*, Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell* **173**, 1535–1548.e1516 (2018).

3. S. Triana *et al.*, Single-cell proteo-genomic reference maps of the hematopoietic system enable the purification and massive profiling of precisely defined cell states. *Nat. Immunol.* **22**, 1577–1589 (2021). doi: 10.1038/s41590-021-01059-0; pmid: 34811546
4. Y. Hao *et al.*, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021). doi: 10.1016/j.cell.2021.04.048; pmid: 34062119
5. M. Lange *et al.*, CellRank for directed single-cell fate mapping. *Nat. Methods* **19**, 159–170 (2022). doi: 10.1038/s41592-021-01346-6; pmid: 35027767
6. P. Weiler, M. Lange, M. Klein, D. Pe'er, F. Theis, CellRank 2: Unified fate mapping in multiview single-cell data. *Nat. Methods* **21**, 1196–1205 (2024). doi: 10.1038/s41592-024-02303-9; pmid: 38871986
7. G. La Manno *et al.*, RNA velocity of single cells. *Nature* **560**, 494–498 (2018). doi: 10.1038/s41586-018-0414-6; pmid: 30089906
8. V. Bergen, M. Lange, S. Peidli, F. A. Wolf, F. J. Theis, Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020). doi: 10.1038/s41587-020-0591-3; pmid: 32747759
9. A. Gayoso *et al.*, Deep generative modeling of transcriptional dynamics for RNA velocity analysis in single cells. *Nat. Methods* **21**, 50–59 (2024). doi: 10.1038/s41592-023-01994-w; pmid: 37735568
10. M. Barile *et al.*, Coordinated changes in gene expression kinetics underlie both mouse and human erythroid maturation. *Genome Biol.* **22**, 197 (2021). doi: 10.1186/s13059-021-02414-y; pmid: 34225769
11. V. Bergen, R. A. Soldatov, P. V. Kharchenko, F. J. Theis, RNA velocity-current challenges and future perspectives. *Mol. Syst. Biol.* **17**, e10282 (2021). doi: 10.15252/msb.202110282; pmid: 34435732
12. Y. Liu, A. Beyer, R. Aebersold, On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**, 535–550 (2016). doi: 10.1016/j.cell.2016.03.014; pmid: 27104977
13. G. Kustatscher *et al.*, Higher-order modular regulation of the human proteome. *Mol. Syst. Biol.* **19**, e9503 (2023). doi: 10.15252/msb.20209503; pmid: 36891684
14. T. Geiger, Tackling tumor complexity with single-cell proteomics. *Nat. Methods* **20**, 324–326 (2023). doi: 10.1038/s41592-023-01784-4; pmid: 36899159
15. C. Buccitelli, M. Selbach, mRNAs, proteins and the emerging principles of gene expression control. *Nat. Rev. Genet.* **21**, 630–644 (2020). doi: 10.1038/s41576-020-0258-4; pmid: 32709985
16. P. R. van den Berg, N. M. L. P. Bérenger-Currias, B. Budnik, N. Slavov, S. Semrau, Integration of a multi-omics stem cell differentiation dataset using a dynamical model. *PLOS Genet.* **19**, e1010744 (2023). doi: 10.1371/journal.pgen.1010744; pmid: 37167320
17. P. Mertins *et al.*, Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534**, 55–62 (2016). doi: 10.1038/nature18003; pmid: 27251275
18. B. Zhang *et al.*, Proteogenomic characterization of human colon and rectal cancer. *Nature* **513**, 382–387 (2014). doi: 10.1038/nature13438; pmid: 25043054
19. L. Peshkin *et al.*, On the Relationship of Protein and mRNA Dynamics in Vertebrate Embryonic Development. *Dev. Cell* **35**, 383–394 (2015). doi: 10.1016/j.devcel.2015.10.010; pmid: 26555057
20. M. Jovanovic *et al.*, Dynamic profiling of the protein life cycle in response to pathogens. *Science* **347**, 1259038 (2015). doi: 10.1126/science.1259038; pmid: 25745177
21. K. Becker *et al.*, Quantifying post-transcriptional regulation in the development of Drosophila melanogaster. *Nat. Commun.* **9**, 4970 (2018). doi: 10.1038/s41467-018-07455-9; pmid: 30478415
22. M. Stoeckius *et al.*, Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017). doi: 10.1038/nmeth.4380; pmid: 28759029
23. A. Konturek-Ciesla *et al.*, Temporal multimodal single-cell profiling of native hematopoiesis illuminates altered differentiation trajectories with age. *Cell Rep.* **42**, 112304 (2023). doi: 10.1016/j.celrep.2023.112304; pmid: 36961818
24. G. Gorin, V. Svensson, L. Pachter, Protein velocity and acceleration from single-cell multiomics experiments. *Genome Biol.* **21**, 39 (2020). doi: 10.1186/s13059-020-1945-3; pmid: 32070398
25. V. Petrosius, E. M. Schoof, Recent advances in the field of single-cell proteomics. *Transl. Oncol.* **27**, 101556 (2023). doi: 10.1016/j.tranon.2022.101556; pmid: 36270102
26. B. Budnik, E. Levy, G. Harmange, N. Slavov, SCoPE-MS: Mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.* **19**, 161 (2018). doi: 10.1186/s13059-018-1547-5; pmid: 30343672
27. E. M. Schoof *et al.*, Quantitative single-cell proteomics as a tool to characterize cellular hierarchies. *Nat. Commun.* **12**, 3341 (2021). doi: 10.1038/s41467-021-23667-y; pmid: 34099695
28. B. Furtwängler *et al.*, Real-Time Search-Assisted Acquisition on a Tribrid Mass Spectrometer Improves Coverage in Multiplexed Single-Cell Proteomics. *Mol. Cell. Proteomics* **21**, 100219 (2022). doi: 10.1016/j.mcpro.2022.100219; pmid: 35219906
29. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018). doi: 10.1186/s13059-017-1382-0; pmid: 29409532
30. C. Vanderaa, L. Gatto, Replication of single-cell proteomics data reveals important computational challenges. *Expert Rev. Proteomics* **18**, 835–843 (2021). doi: 10.1080/14789450.2021.1988571; pmid: 34602016
31. E. Becht *et al.*, Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–44 (2018). doi: 10.1038/nbt.4314; pmid: 30531897
32. M. D. Luecken *et al.*, Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods* **19**, 41–50 (2022). doi: 10.1038/s41592-021-01336-8; pmid: 34949812
33. N. A. Jakobsen *et al.*, Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging. *Cell Stem Cell* **31**, 1127–1144.e17 (2024). doi: 10.1016/j.stem.2024.05.010; pmid: 38917807
34. F. Notta *et al.*, Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science* **351**, aab2116 (2016). doi: 10.1126/science.aab2116; pmid: 26541609
35. F. Paul *et al.*, Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell* **163**, 1663–1677 (2015). doi: 10.1016/j.cell.2015.11.013; pmid: 26627738
36. L. Perié, K. R. Duffy, L. Kok, R. J. de Boer, T. N. Schumacher, The Branching Point in Erythro-Myeloid Differentiation. *Cell* **163**, 1655–1662 (2015). doi: 10.1016/j.cell.2015.11.059; pmid: 26687356
37. Y. Mori *et al.*, Identification of the human eosinophil lineage-committed progenitor: Revision of phenotypic definition of the human common myeloid progenitor. *J. Exp. Med.* **206**, 183–193 (2009). doi: 10.1084/jem.20081756; pmid: 19114669
38. R. Drissen, S. Thongjuea, K. Theilgaard-Mönch, C. Nerlov, Identification of two distinct pathways of human myelopoiesis. *Sci. Immunol.* **4**, eaau7148 (2019). doi: 10.1126/sciimmunol.aau7148; pmid: 31126997
39. J. Grootens *et al.*, Single-cell analysis reveals the KIT D816V mutation in haematopoietic stem and progenitor cells in systemic mastocytosis. *EBioMedicine* **43**, 150–158 (2019). doi: 10.1016/j.ebiom.2019.03.089; pmid: 30975542
40. K. Reckzeh *et al.*, Human adult HSCs can be discriminated from lineage-committed HPCs by the expression of endomucin. *Blood Adv.* **2**, 1628–1632 (2018). doi: 10.1182/bloodadvances.2018015743; pmid: 29986855
41. C. Sanada *et al.*, Adult human megakaryocyte-erythroid progenitors are in the CD34+CD38mid fraction. *Blood* **128**, 923–933 (2016). doi: 10.1182/blood-2016-01-693705; pmid: 27268089
42. S. Amon *et al.*, Sensitive Quantitative Proteomics of Human Hematopoietic Stem and Progenitor Cells by Data-independent Acquisition Mass Spectrometry. *Mol. Cell. Proteomics* **18**, 1454–1467 (2019). doi: 10.1074/mcp.TIR119.001431; pmid: 30975897
43. O. Valiron, C. Gorka, Histone H1(0) expression is restricted to progenitor cells during human hematopoiesis. *Eur. J. Cell Biol.* **72**, 39–45 (1997). pmid: 9013724
44. J. L. Villeval *et al.*, Carbonic anhydrase I is an early specific marker of normal human erythroid differentiation. *Blood* **66**, 1162–1170 (1985). doi: 10.1182/blood.V66.5.1162.1162; pmid: 3931725
45. D. D. Metcalfe *et al.*, Biomarkers of the involvement of mast cells, basophils and eosinophils in asthma and allergic diseases. *World Allergy Organ. J.* **9**, 7 (2016). doi: 10.1186/s40413-016-0094-3; pmid: 26904159
46. S. Rørvig, O. Østergaard, N. H. Heegaard, N. Borregaard, Proteome profiling of human neutrophil granule subsets, secretory vesicles, and cell membrane: Correlation with transcriptome profiling of neutrophil precursors. *J. Leukoc. Biol.* **94**, 711–721 (2013). doi: 10.1189/jlb.1212619; pmid: 23650620
47. Tabula Muris Consortium; Overall coordination; Logistical coordination; Organ collection and processing; Library preparation and sequencing; Computational data analysis; Cell type annotation; Writing group; Supplemental text writing group; Principal investigators, Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018). doi: 10.1038/s41586-018-0590-4; pmid: 30283141
48. D. A. Anderson III, C. A. Dutertre, F. Ginhoux, K. M. Murphy, Genetic models of human and mouse dendritic cell development and function. *Nat. Rev. Immunol.* **21**, 101–115 (2021). doi: 10.1038/s41577-020-00413-x; pmid: 32908299
49. L. Broderick *et al.*, Mutations in topoisomerase IIβ result in a B cell immunodeficiency. *Nat. Commun.* **10**, 3644 (2019). doi: 10.1038/s41467-019-11570-6; pmid: 31409799
50. Z. Fang, X. Liu, G. Peltz, GSEApy: A comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* **39**, btac757 (2023). doi: 10.1093/bioinformatics/btac757; pmid: 36426870
51. T. Simsek *et al.*, The distinct metabolic profile of hematopoietic stem cells reflects their location in a hypoxic niche. *Cell Stem Cell* **7**, 380–390 (2010). doi: 10.1016/j.stem.2010.07.011; pmid: 20804973
52. A. Gayoso *et al.*, Joint probabilistic modeling of single-cell multi-omic data with totalVI. *Nat. Methods* **18**, 272–282 (2021). doi: 10.1038/s41592-020-01050-x; pmid: 33589839
53. R. Lopez, J. Regier, M. B. Cole, M. I. Jordan, N. Yosef, Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018). doi: 10.1038/s41592-018-0229-2; pmid: 30504886
54. F. A. Wolf *et al.*, PAGA: Graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59 (2019). doi: 10.1186/s13059-019-1663-x; pmid: 30890159
55. N. Cabezas-Wallscheid *et al.*, Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. *Cell Stem Cell* **15**, 507–522 (2014). doi: 10.1016/j.stem.2014.07.005; pmid: 25158935
56. Z. J. Cao, G. Gao, Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nat. Biotechnol.* **40**, 1458–1466 (2022). doi: 10.1038/s41587-022-01284-4; pmid: 35501393

57. Y. Liu, Y. Chen, X. Deng, J. Zhou, ATF3 Prevents Stress-Induced Hematopoietic Stem Cell Exhaustion. *Front. Cell Dev. Biol.* **8**, 585771 (2020). doi: 10.3389/fcell.2020.585771; pmid: 33195236
58. J. Jiang *et al.*, A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nat. Cell Biol.* **10**, 353–360 (2008). doi: 10.1038/ncb1698; pmid: 18264089
59. E. Laurenti *et al.*, CDK6 levels regulate quiescence exit in human hematopoietic stem cells. *Cell Stem Cell* **16**, 302–313 (2015). doi: 10.1016/j.stem.2015.01.017; pmid: 25704240
60. H. Zhang, M. Ye, R. S. Welner, D. G. Tenen, Sox4 Is Required for the Formation and Maintenance of Multipotent Progenitors. *Blood* **124**, 1577 (2014). doi: 10.1182/blood.V124.21.1577.1577
61. S. Zaunz *et al.*, Loss of the Base Excision Repair Gene Leads to Dysfunctional Adult Hematopoietic Stem and Progenitor Cells. *Blood* **138**, 3267 (2021). doi: 10.1182/blood-2021-146369
62. A. Subramanian *et al.*, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005). doi: 10.1073/pnas.0506580102; pmid: 16199517
63. P. Hernández-Malmierca *et al.*, Antigen presentation safeguards the integrity of the hematopoietic stem cell pool. *Cell Stem Cell* **29**, 760–775.e10 (2022). doi: 10.1016/j.stem.2022.04.007; pmid: 35523139
64. J. Hoggatt, P. Singh, J. Sampath, L. M. Pelus, Prostaglandin E2 enhances hematopoietic stem cell homing, survival, and proliferation. *Blood* **113**, 5444–5455 (2009). doi: 10.1182/blood-2009-01-201335; pmid: 19324903
65. T. Michurina *et al.*, Nitric oxide is a regulator of hematopoietic stem cell activity. *Mol. Ther.* **10**, 241–248 (2004). doi: 10.1016/j.ymthe.2004.05.030; pmid: 15294171
66. K. Banki, E. Hutter, E. Colombo, N. J. Gonchoroff, A. Perl, Glutathione levels and sensitivity to apoptosis are regulated by changes in transaldolase expression. *J. Biol. Chem.* **271**, 32994–33001 (1996). doi: 10.1074/jbc.271.51.32994; pmid: 8955144
67. A. Almars, "Heterochromatin protein HP1BP3 modulates haematopoiesis and leukemogenesis", thesis, University of Nottingham (2021).
68. O. Bereshchenko *et al.*, Deficiency and haploinsufficiency of histone macroH2A1.1 in mice recapitulate hematopoietic defects of human myelodysplastic syndrome. *Clin. Epigenetics* **11**, 121 (2019). doi: 10.1186/s13148-019-0724-z; pmid: 31439048
69. Z. Thompson *et al.*, High Mobility Group A1 Chromatin Regulators Drive Regenerative Capacity during Stress Hematopoiesis through Modulation of TNF-α and Aging Transcriptional Networks. *Blood* **142**, 4050 (2023). doi: 10.1182/blood-2023-189964
70. J. W. Shin *et al.*, Lamins regulate cell trafficking and lineage maturation of adult human hematopoietic cells. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 18892–18897 (2013). doi: 10.1073/pnas.1304996110; pmid: 24191023
71. Y. Hu, G. K. Smyth, ELDA: Extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J. Immunol. Methods* **347**, 70–78 (2009). doi: 10.1016/j.jim.2009.06.008; pmid: 19567251
72. S. Hadjur *et al.*, Defective hematopoiesis and hepatic steatosis in mice with combined deficiencies of the genes encoding Fancc and Cu/Zn superoxide dismutase. *Blood* **98**, 1003–1011 (2001). doi: 10.1182/blood.V98.4.1003; pmid: 11493445
73. J. S. Friedman *et al.*, Absence of mitochondrial superoxide dismutase results in a murine hemolytic anemia responsive to therapy with a catalytic antioxidant. *J. Exp. Med.* **193**, 925–934 (2001). doi: 10.1084/jem.193.8.925; pmid: 11304553
74. G. Kustatscher, P. Grabowski, J. Rappsilber, Pervasive coexpression of spatially proximal genes is buffered at the protein level. *Mol. Syst. Biol.* **13**, 937 (2017). doi: 10.15252/msb.20177548; pmid: 28835372
75. J. K. Pickrell *et al.*, Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**, 768–772 (2010). doi: 10.1038/nature08872; pmid: 20220758
76. A. Battle *et al.*, Impact of regulatory variation from RNA to protein. *Science* **347**, 664–667 (2015). doi: 10.1126/science.1260793; pmid: 25657249
77. E. McShane *et al.*, Kinetic Analysis of Protein Stability Reveals Age-Dependent Degradation. *Cell* **167**, 803–815.e21 (2016). doi: 10.1016/j.cell.2016.09.015; pmid: 27720452
78. D. P. Zolg *et al.*, INFERYS rescoring: Boosting peptide identifications and scoring confidence of database search results. *Rapid Commun. Mass Spectrom.* **39**, 9128 (2021). doi: 10.1002/rcm.9128; pmid: 34015160
79. S. White *et al.*, FlowKit: A Python Toolkit for Integrated Manual and Automated Cytometry Analysis Workflows. *Front. Immunol.* **12**, 768541 (2021). doi: 10.3389/fimmu.2021.768541; pmid: 34804056
80. W. E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007). doi: 10.1093/biostatistics/kxj037; pmid: 16632515
81. F. Supek, M. Bošnjak, N. Škunca, T. Šmuc, REVIGO summarizes and visualizes long lists of gene ontology terms. *PLOS ONE* **6**, e21800 (2011). doi: 10.1371/journal.pone.0021800; pmid: 21789182
82. B. Furtwängler, N. Üresin, S. Richter, F. Theis, E. M. Schoof, B. T. Porse, Mapping early human blood cell differentiation using single-cell proteomics and transcriptomics, Version 1, Zenodo (2025); https://zenodo.org/records/15554000.

ACKNOWLEDGMENTS

We thank P. Weiler for the setup of the first version of scProtVelo and the FACS core facility at BRIC for assistance with cell sorting. We thank A.-K. Frank for coordinating access to human samples. **Funding:** B.F. is the recipient of a Copenhagen Bioscience Ph.D. stipend by the Novo Nordisk Foundation, NNF19SA0035442. S.R. is supported by the Helmholtz Association under the joint research school Munich School for Data Science. Work in the Porse lab was supported by grants from the Svend Andersen Foundation, the Candys Foundation, the Danish Cancer Society, the Eva and Henry Frænkel Memorial Foundation, the Independent Research Fund Denmark, and through a center grant from the Novo Nordisk Foundation (Novo Nordisk Foundation Center for Stem Cell Biology, DanStem; grant NNF17CC0027852). This work has been performed in the context of the Danish Research Center for Precision Medicine in Blood Cancers funded by the Danish Cancer Society (R223-A13071) and Greater Copenhagen Health Science Partners. Work in the Schoof lab was supported by grants from the Independent Research Fund Denmark (2067-00053B), Danish Cancer Society (R324-A17978), Novo Nordisk Foundation (NNF21OC0071016), and the Lundbeck Foundation (R413-2022-869). **Author contributions:** N.Ü. performed scp-MS sample collection and preparation. B.F. performed scp-MS measurements. D.B. performed CITE-seq sample preparation and measurement. A.W. performed CITE-seq raw data analysis. M.B.S. performed the CRISPR knockout experiments. S.R. and F.T. created the scProtVelo modeling. H.H. contributed to data interpretation. B.F. and S.R. analyzed and interpreted the data. K.G. and K.T.-M. aspirated human bone marrow samples, provided methodology, and contributed to data interpretation. Supervision: F.T., E.S., and B.T.P. E.S. and B.T.P. conceived the study. Writing - original draft: B.F., N.Ü., and S.R. Writing - review & editing: F.T., E.S., and B.T.P. All authors approved the final version of the paper. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The raw scp-MS and scRNA-seq data generated in this study have been deposited in European Genome-phenome Archive (EGA) under accession number EGAS00001007930. Code and data, including experimental data and scProtVelo, are available at Zenodo (*82*). The scProtVelo package is also available on github via github.com/theislab/scProtVelo. The external HSPC proteomics data were accessed via the PRIDE repository identifier PXD009255, and the accompanying RNA sequencing data were accessed via the Expression Omnibus with identifier GSE113182. **License information:** Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/content/page/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adr8785
Supplementary Text; Figs. S1 to S25; Tables S1 to S3; References (*83*–*87*);
MDAR Reproducibility Checklist

Submitted 19 July 2024; resubmitted 15 February 2025; accepted 14 July 2025; published online 21 August 2025

10.1126/science.adr8785

RESEARCH ARTICLE SUMMARY

LINEAGE TRACING

High-resolution spatial mapping of cell state and lineage dynamics in vivo with PEtracer

Luke W. Koblan*†, Kathryn E. Yost*†, Pu Zheng†, William N. Colgan†, Matthew G. Jones, Dian Yang, Arhan Kumar, Jaspreet Sandhu, Alexandra Schnell, Dawei Sun, Can Ergen, Reuben A. Saunders, Xiaowei Zhuang, William E. Allen, Nir Yosef, Jonathan S. Weissman*

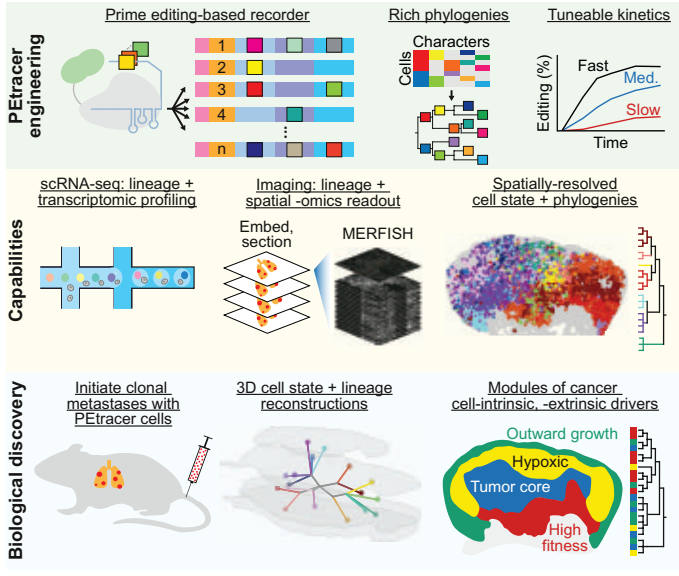
Full article and list of author affiliations: https://doi.org/10.1126/science.adx3800

INTRODUCTION: Organismal lineage maps, such as that of *Caenorhabditis elegans*, have provided critical insights into the dynamic interplay of cell state, ontogeny, and spatial cues that coordinate cell fate determination. However, creating analogous, comprehensive maps in mammals has remained out of reach. The ability to jointly track cell state and lineage with high spatiotemporal resolution would offer powerful insights into complex biological processes in development and disease.

RATIONALE: We sought to develop an engineered lineage tracing system that enables joint cell state and lineage profiling in vivo across relevant spatial and temporal scales—from single cells to whole tissues, and ranging from a single cell division to the life of an organism. Imaging-based cell profiling methods, such as multiplexed error-robust fluorescence in situ hybridization (MERFISH), are well suited for multimodal cell phenotyping at tissue scale with subcellular resolution. Thus, the design and engineering of our lineage tracing system were guided by compatibility with imaging readouts. Specifically, using FISH-based imaging to assay lineage information requires a predefined set of lineage marks that are distinguishable by hybridization probes. We selected prime editing as the basis for an evolving lineage recorder capable of continuously installing immutable, predefined marks with sufficient diversity to enable robust lineage reconstruction.

RESULTS: Guided by extensive theoretical modeling and experimental considerations, we built and optimized a prime editing (PE)–based lineage tracing system (PEtracer), which installs one of eight five-nucleotide marks at dozens of genomically integrated, barcoded lineage tracing cassettes. We tuned editing kinetics to record lineage information over diverse time frames ranging from days to months. Using cells containing predefined lineage information, we confirmed high readout efficiency and accuracy by both sequencing and imaging. Furthermore, rigorous benchmarking experiments demonstrated near-perfect phylogenetic reconstruction accuracy, establishing a strong foundation for in vivo studies.

We then applied PEtracer to a transplantable, syngeneic mouse model of breast cancer lung metastasis to study the dynamics of cancer cell extravasation, colonization, and outgrowth. By combining high-resolution, MERFISH-based spatial transcriptomic profiling with rich phylogenies for tens of thousands of cells in clonally seeded tumors, we provide an unprecedented view of cancer growth in space and time as well as the organization of immune and stromal cells within these tumors. Joint analysis of cell state, lineage, and spatial microenvironmental factors across multiple sections of sampled tumors revealed malignant cell modules that link cell fitness and plasticity to both cell-intrinsic and -extrinsic drivers. Specifically, we identified a niche adjacent to the normal lung associated with a heritable, high-fitness, and epithelial-like state.



PEtracer enables spatially resolved cell state profiling and lineage tracing in vivo. (Top) A prime editing–based recorder with engineered barcodes and tunable editing kinetics generates rich phylogenetic data. (Middle) Joint single-cell sequencing and MERFISH imaging provide multimodal and spatially resolved cell state and lineage information. (Bottom) Applied to metastatic tumors, PEtracer reconstructs three-dimensional (3D) clonal growth, revealing distinct cell modules and spatial fitness programs.

CONCLUSION: PEtracer is an evolving lineage tracing system for high-resolution cell state and lineage profiling across space and time in vivo. Our system integrates cutting-edge approaches in genome engineering, lineage tracing, and advanced imaging methods and can be leveraged to study models of cancer initiation, progression, metastasis, and drug resistance as well as fundamental questions in development, stem cell biology, and homeostasis. Moreover, PEtracer is positioned to generate spatially integrated multimodal datasets to train burgeoning artificial intelligence foundation models of dynamic cellular behaviors in health and disease. □

*Corresponding author. Email: lkoblan@wi.mit.edu (L.W.K.); kyost@wi.mit.edu (K.E.Y.); weissman@wi.mit.edu (J.S.W.) †These authors contributed equally to this work. Cite this article as L. W. Koblan *et al.*, *Science* 390, eadx3800 (2025). DOI: 10.1126/science.adx3800

LINEAGE TRACING

High-resolution spatial mapping of cell state and lineage dynamics in vivo with PEtracer

Luke W. Koblan[1,2,3]*†, Kathryn E. Yost[1,2,3]*†, Pu Zheng[1,2,3]†, William N. Colgan[1,2,3]†, Matthew G. Jones[4]‡§¶, Dian Yang[5,6], Arhan Kumar[1,2,3], Jaspreet Sandhu[1,2,3,7], Alexandra Schnell[1,2,3], Dawei Sun[8,9], Can Ergen[10,11], Reuben A. Saunders[1,2,3,12], Xiaowei Zhuang[2,13,14], William E. Allen[2,12]#, Nir Yosef[15], Jonathan S. Weissman[1,2,3,16]*

Charting the spatiotemporal dynamics of cell fate determination in development and disease is a long-standing objective in biology. Here, we present the design, development, and extensive validation of PEtracer, a prime editing (PE)–based, evolving lineage tracing technology compatible with both single-cell sequencing and multimodal imaging methodologies, created to jointly profile cell state and lineage in dissociated cells or while preserving cellular context in tissues with high spatial resolution. Using PEtracer coupled with MERFISH spatial transcriptomic profiling in a syngeneic mouse model of tumor metastasis, we reconstructed the growth of individually seeded tumors in vivo and uncovered distinct modules of cell-intrinsic and -extrinsic factors that coordinate tumor growth. More generally, PEtracer enables systematic characterization of cell state and lineage relationships in intact tissues over biologically relevant temporal and spatial scales.

Defining the history of cell division, differentiation, movement, and death provides fundamental insights into how complex structures emerge in multicellular life (*1*–*3*), including embryonic development (*4*–*7*), tissue maintenance by stem cells (*8*–*12*), disease progression, and tumor evolution (*13*–*20*). Accordingly, many approaches have been developed to track cell lineages with naturally occurring mutations or engineered systems (*21*–*32*).

Initial prospective lineage tracing efforts and modern elaborations track static marks in cells—including dyes, fluorescent proteins, or recombined DNA reporter sequences—that provide views of clonal tissue architecture (*21*, *33*–*38*). Although these methods can preserve spatial context, they typically resolve progenitor outputs from predefined time points, limiting their capacity to capture dynamic cellular processes over time.

Recently engineered molecular recording technologies that pair cell lineage tracing with rich phenotyping have greatly expanded our ability to construct cellular phylogenies with high temporal resolution (*39*–*52*). Typically, these approaches stochastically install heritable lineage marks (LMs) at predetermined sites, for example using Cas9-based genome editing, which can then be read out along with cell state using single-cell approaches. Systems with continuous accrual of heritable marks and sufficient diversity can enable reconstruction of vast phylogenetic trees, ideally recording the complete lineage history of all cells in a tissue (*39*, *43*, *44*, *47*, *53*, *54*). There is growing interest in developing spatially resolved lineage readouts (*34*, *40*, *42*, *55*, *56*) that leverage imaging-based cell profiling to capture multimodal cellular information with true single-cell spatial resolution at tissue scale (*57*–*65*). However, integrating the depth and timescale of evolving lineage tracing with high-dimensional, spatially resolved cell-state mapping in vivo remains challenging.

Here, we describe the design, engineering, and rigorous validation of PEtracer, a prime editing (PE) (*66*–*69*)–based lineage tracing system (Fig. 1, A to C). We use PE to stochastically install immutable, predefined five-nucleotide (nt) LMs that can be robustly assayed by either single-cell sequencing for whole-transcriptome profiling (*39*, *53*–*55*) [e.g., droplet-based single-cell RNA sequencing (scRNA-seq)] or fluorescence in situ hybridization (FISH)–based imaging—specifically, multiplexed error-robust FISH (MERFISH)—for spatial transcriptomic profiling at subcellular resolution (*56*, *58*). This readout flexibility enables both transcriptome-wide profiling and high-resolution spatial imaging. We applied PEtracer to study the 4T1 transplantable model of murine metastatic breast cancer and reconstructed the three-dimensional (3D) growth of individually seeded tumors in vivo to dissect how cell-intrinsic and -extrinsic factors coordinate cellular behaviors during metastatic outgrowth. The PEtracer system represents an important step toward creating comprehensive cell-fate maps of development and disease.

In silico modeling of lineage tracing systems

Compatibility with imaging readouts guided the design and engineering of our lineage tracing system. In particular, assaying lineage information by using FISH-based imaging requires that LMs be discriminable with a predefined probe set. Thus, imaging readouts of current Cas9-based lineage tracers would be challenging because these systems install hundreds of varied LMs (*39*, *53*, *54*, *70*, *71*). We therefore used in silico simulations to assess whether fewer LMs could still yield high-quality phylogenies and to evaluate how key, experimentally relevant parameters influence reconstruction performance (*72*).

We systematically varied the number of cells being traced, the number of LMs installed at each edit site (ES), the relative installation efficiencies for different LMs, the number of distinct ESs used for tree building, the fraction of ESs with a LM, and LM detection efficiency. Using simulated phylogenies, we quantified performance by calculating both the Robinson-Foulds (RF) distance (*73*) and the mean number of triplets shared (*74*) between the reconstructed and ground-truth phylogenies (Materials and methods).

Reconstruction accuracy improved with both higher numbers of LMs and more balanced LM distributions (i.e., higher normalized entropy; H_{norm}) (Fig. 1D and table S1) (*75*). However, the incremental benefit of additional LMs diminishes as their number increases (Fig. 1D). Eight well-balanced ($H_{\mathrm{norm}} > 0.9$) LMs provide a good compromise between reconstruction accuracy and the feasibility of FISH-based imaging readouts (fig. S1, A to F, and tables S2 and S3) (*76*–*78*). As expected, low detection efficiency can degrade reconstruction performance, but this effect can be counteracted by increasing ES number (fig. S1E). Overall, our simulations demonstrate that a system with eight well-balanced LMs performs equivalently to existing Cas9-based tracers, reconstructing accurate phylogenies across a range of experimental parameters in a manner robust to imperfect detection rates (fig. S1E).

[1]Whitehead Institute for Biomedical Research, Cambridge, MA, USA. [2]Howard Hughes Medical Institute, Chevy Chase, MD, USA. [3]David H. Koch Center for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, USA. [4]Department of Dermatology, Stanford University, Stanford, CA, USA. [5]Department of Systems Biology, Columbia University, New York, NY, USA. [6]Department of Molecular Pharmacology and Therapeutics, Columbia University, New York, NY, USA. [7]Division of Gastroenterology, Massachusetts General Hospital, Boston, MA, USA. [8]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [9]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA. [10]Center for Computational Biology, University of California, Berkeley, Berkeley, CA, USA. [11]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA, USA. [12]Harvard Society of Fellows, Harvard University, Cambridge, MA, USA. [13]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA. [14]Department of Physics, Harvard University, Cambridge, MA, USA. [15]Department of Systems Immunology, Weizmann Institute of Science, Rehovot, Israel. [16]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. ***Corresponding author. Email: lkoblan@wi.mit.edu (L.W.K.); kyost@wi.mit.edu (K.E.Y.); weissman@wi.mit.edu (J.S.W.)** †These authors contributed equally to this work. ‡Present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. §Present address: David H. Koch Center for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, USA. ¶Present address: Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, USA. #Present address: Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA, USA.

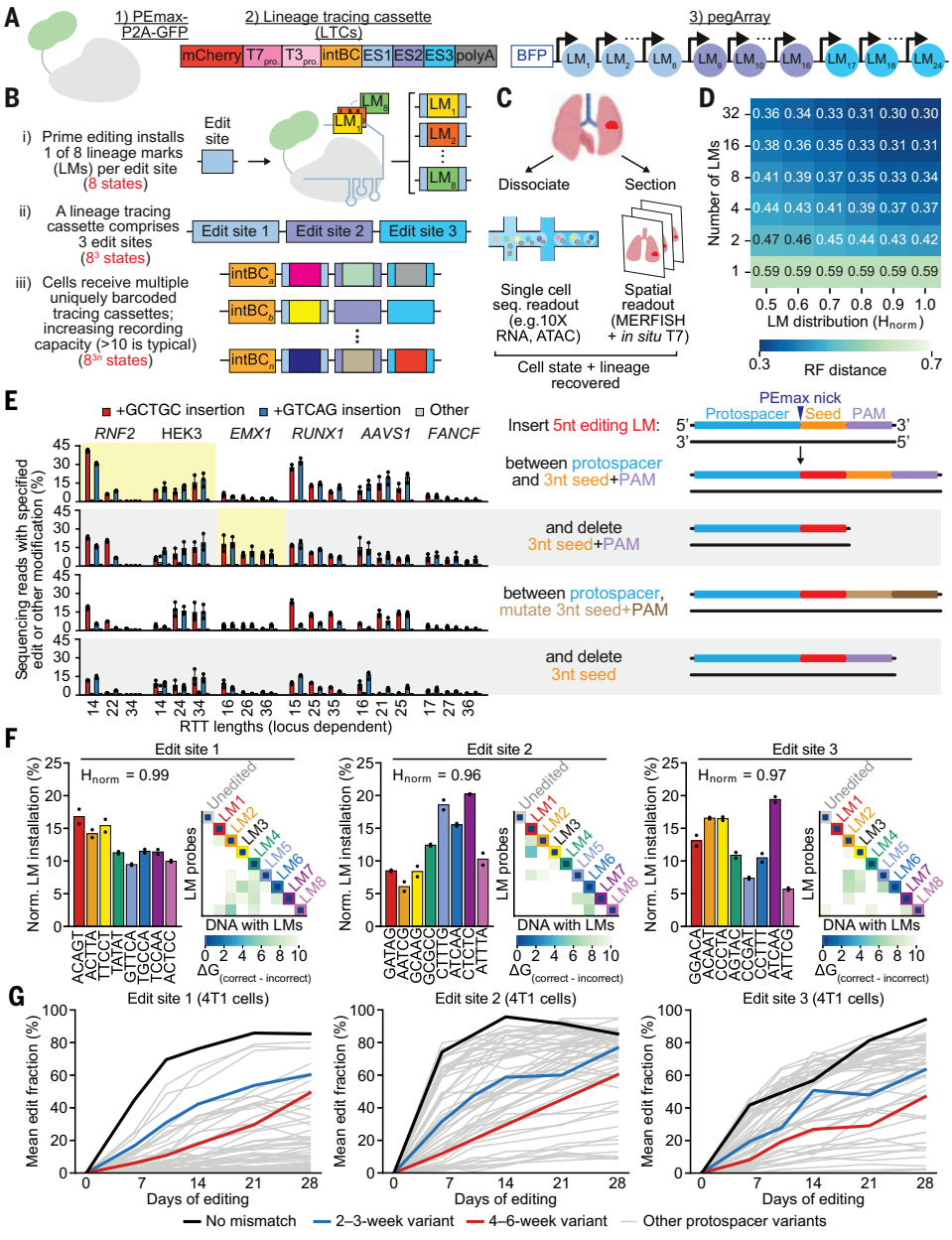


Fig. 1. Design and engineering of the PEtracer system. (**A**) Components of the PEtracer system: 1) the PEmax-P2A-GFP genome editing agent; 2) lineage recording cassettes (LTCs) in the 3′ UTR of an *mCherry* gene, each cassette containing a distinct integration barcode (intBC) and three distinct edit sites (ESs) that can be assayed either by droplet-based single-cell sequencing or by in situ transcription through T7 or T3 polymerization; and 3) prime editing guide RNA arrays (pegArrays), which encode all 24 lineage marks (LMs) used for lineage tracing (eight LMs for each of the three ESs). (**B**) Schematic of evolving lineage tracing using PEtracer. PEmax installs LMs at predetermined ESs, each LTC contains three ESs, and intBCs enumerate distinct integrated cassettes. (**C**) PEtracer cell state and lineage readout compatibility with both dissociative single-cell readouts and those that preserve the architecture of the original tissue context. Lung image from BioRender.com. (**D**) Mean Robinson-Foulds (RF) reconstruction error for simulated phylogenies ($n = 10$) while varying the number of LMs and normalized entropy (H_{norm}) of their relative installation efficiencies. (**E**) Optimization of LM insertion strategies (diagrams on right) across six endogenous genomic loci (*RNF2*, HEK3, *EMX1*, *RUNX1*, *AAVS1*, and *FANCF*) using two representative 5-nt LMs in HEK293T cells. Four strategies for installing 5-nt LMs at ESs were tested: 1) inserting a 5-nt mark between the protospacer adjacent motif (PAM) and the adjacent genomic target sequence recognized by the protospacer of our pegRNAs, 2) deleting the PAM and 3 nt of the genomic target seed region [positions 18 to 20 of the protospacer, where the PAM NGG (where N is any nucleotide) are positions 21 to 23] and replacing these deleted bases with a 5-nt mark, 3) performing the same edit as variant 1 but recoding the 3-nt seed and PAM, and 4) replacing the 3-nt seed sequence with a 5-nt mark. Editing rates reflect the percentage of sequencing reads that contain the intended edit out of all aligned reads. Mean of three biological replicates ± SD is depicted. Sites and strategies used for subsequent experiments are highlighted in yellow. Reverse transcription template (RTT) lengths are specific to each ES. (**F**) Bar plot of the normalized (Norm.) LM installation efficiencies for the eight LMs in the final 24-mer pegArray (without protospacer mismatches) at each of the three finalized ESs in B16-F10 cells (on left for each ES) averaged across two biological replicates; H_{norm} values are provided for each ES. Predicted probe hybridization specificity (Gibbs free energy = ΔG) for discriminating each of the eight LMs from one another (on right for each site). (**G**) Editing kinetics for protospacer mismatch variants at ESs 1, 2, and 3 in 4T1 cells. Gray lines show all protospacer variants tested; no mismatch protospacers shown in black; protospacer mismatches selected for 2- to 3-week timescales shown in blue; 4- to 6-week protospacer mismatches in red.

There is a tradeoff between LM installation rate and the ability to mark later cell divisions because unmodified sites diminish over time (fig. S1, G and H). Our simulations indicate that 60 to 80% edit saturation maximizes reconstruction accuracy, highlighting the need to precisely tune editing kinetics for different biological timescales (fig. S1, E and H). Because tracing longer timescales requires slower editing rates, more ESs are needed to maintain lineage resolution comparable to that of shorter timescales given similar cell division rates. Our simulations indicate that with optimized kinetics, 75 ESs can theoretically mark 90% of divisions in a 10^6-cell phylogeny generated by 20 doublings, and 100 ESs can achieve the same resolution for 10^9 cells from 30 doublings (fig. S1I and table S4).

Design of the PEtracer system

PE is a gene editing technology that is well suited to satisfy the requirements established in our simulation work because it enables precise insertion of diverse, predefined sequences in genomic DNA with minimal undesired by-products or editing-induced toxicity (*66*, *79*–*83*). A predefined set of LMs can be encoded by distinct prime editing guide RNAs (pegRNAs) targeting the same ES, and high H_{norm} values can be achieved by selecting pegRNAs that evenly compete during LM installation. Furthermore, PE-installed LMs can be designed to precisely alter the protospacer adjacent motif (PAM) or protospacer seed region to prevent re-editing, thus installing predefined and immutable marks that can be read out with FISH-based imaging.

We designed a PE-based lineage tracing system (PEtracer) with three components: (i) the PEmax editor (*67*), whose expression can be monitored using coexpressed green fluorescent protein (GFP); (ii) lineage tracing cassettes (LTCs) in the 3′ untranslated region (UTR) of an *mCherry* gene, containing a variable DNA integration barcode (intBC) and three distinct ESs where LMs can be continuously and stochastically installed; and (iii) pegRNA arrays (pegArrays) encoding 24 LMs (8 LMs per ES) linked to a blue fluorescent protein (BFP) gene where each pegRNA encodes a distinct LM (Fig. 1, A to C). We selected PE2-type editing as opposed to PE3-type, which involves opposite strand nicking, to avoid double-stranded DNA break–derived insertions and deletions (indels) (*66*). intBCs and LMs can be read out by sequencing of polyadenylated transcripts or imaging of in situ transcribed LTCs detected with FISH probes (Fig. 1C). Because LTCs are randomly integrated throughout the genome, they are spatially distinguishable and can be decoded with hybridization probes after T7 transcription in situ (*56*). This system is compatible with endogenous RNA MERFISH, enabling simultaneous readout of lineage and gene expression in tissue sections.

To align PEtracer performance with benchmarks established with our simulations, we performed three core technical optimizations to our system: (i) identification of ESs and pegRNAs that enable rapid irreversible LM installation, (ii) selection of predefined LMs installed with similar efficiencies and maximally discriminable by hybridization-based approaches, and (iii) tuning of LM installation kinetics to use the PEtracer system across a broad range of biological timescales.

PEtracer ES and LM installation modality optimizations

We first determined the ES sequences, DNA editing strategy, and LM lengths that allow for highly efficient editing and accurate detection by both sequencing and imaging readouts. We reasoned that 5-nt marks would be installed with high efficiency by PE2 (*84*); would evade efficient mismatch repair detection (*67*), which can limit PE editing efficiencies in a cell type–specific manner; and would provide diverse sequences for robust discrimination with FISH probes (fig. S1J, table S5, and Materials and methods).

We tested four 5-nt LM installation strategies at candidate ESs (Fig. 1E, right), each designed to prevent re-editing to both boost PE efficiency and ensure immutable, heritable marking (*66*). We installed two representative LMs, utilizing these editing strategies across six genomic loci using pegRNAs with fixed primer binding site lengths and a range of reverse transcription template (RTT) lengths (*85*). We selected three loci with consistent and robust editing [Fig. 1E; *RNF2*, HEK293 site 3 (HEK3), and *EMX1*] and further improved editing efficiency at these sites 1.5- to 3.6-fold through RTT optimization, using engineered pegRNAs (epegRNAs) (fig. S2A) (*85*). Notably, our initial optimizations were performed at human genomic loci; thus our pegRNAs would modify both endogenous sequences in human cells and LTC ESs. We therefore generated orthogonalized ES and epegRNA variants that maintain high editing efficiency while preventing editing of endogenous loci (fig. S2, B to D, and Materials and methods) (*86*).

As has been done previously (*39*), we constructed LTCs by inserting ES sequences in the 3′ UTR of an mCherry reporter to enable fluorescence-activated cell sorting (FACS) enrichment of cells with high numbers of LTC integrations (Fig. 1A). The orthogonalized variants of the *RNF2*, HEK3, and *EMX1* loci were termed ES1, ES2, and ES3, respectively, with all editing occurring on the same DNA strand to minimize PE-mediated generation of tandem duplications or indels (*67*, *87*). LTCs are highly expressed for efficient capture by scRNA-seq. To facilitate detection by imaging, we introduced T7 and T3 promoters in the LTC to enable in situ transcription and thus amplification of integrated cassettes (*56*). Each cassette is marked by an intBC containing both a sequencing barcode (30 nt) and a FISH based–imaging barcode (183 nt). We generated a library of 2171 intBC-tagged LTCs (annotated as intBC 1 to 2171; the whitelist intBC number was randomly assigned) to enable polyclonal cell-line engineering in which individual clones can be distinguished by their set of intBCs (table S6).

Selection of balanced LM sequences compatible with imaging-based readout

With LTCs and preliminary epegRNAs designed, we next sought to identify sets of eight 5-nt LMs for each ES that were both efficiently installed and maximally discriminable by hybridization-based readouts. We tested all 1024 possible LMs for each target locus and identified ES-specific LMs that were installed with high efficiency (figs. S2E and S3A and table S7). Among this efficiently installed set, we computed how effectively each LM could be discriminated by hybridization probes (fig. S3, A and B; table S8; and Materials and methods). For each ES, we then nominated 20 5-nt LMs and, on the basis of retesting at orthogonalized ESs, selected the best-performing 18 for subsequent engineering efforts (fig. S3, A to C).

Lastly, to ensure that each LM was installed with a similar efficiency when all epegRNAs are expressed in the same cell, we created arrays comprising eight epegRNAs each (8-mer arrays, table S9). Initial 8-mer pegArrays showed reasonably balanced performance ($0.90<H_{norm}<0.98$; fig. S4A), which we further optimized (H_{norm} for ES1: 0.98; ES2: 0.99; ES3: 0.96; fig. S4B) to construct our final 24-mer array consisting of an 8-mer for each ES (Fig. 1F). The final sets of eight LMs for each ES are predicted to be highly discriminable from one another and the unedited state, according to calculated Gibbs free-energy differences, ΔG (Fig. 1F and table S8).

Kinetic tuning of epegRNAs to use PEtracer across diverse timescales

To identify epegRNAs that enable continuous editing over timescales ranging from several days to months, we tested a library of mismatches in the protospacers of epegRNAs targeting each ES (*39*). We exhaustively altered all protospacer positions except the first G for epegRNAs targeting each ES (57 per ES, Fig. 1G) and tested them with scRNA-seq–compatible lentiviral libraries (*88*) (Materials and methods) in two DNA mismatch repair–competent mouse cancer cell lines used to model metastasis (B16-F10 and 4T1) (*89*–*91*).

We identified numerous protospacer mismatches with slower LM installation rates than those of unmodified protospacer sequences (Fig. 1G and fig. S4C). To estimate the rate at which ESs are modified (edits per ES per day), we fit curves modeling the exponential decay of unmodified

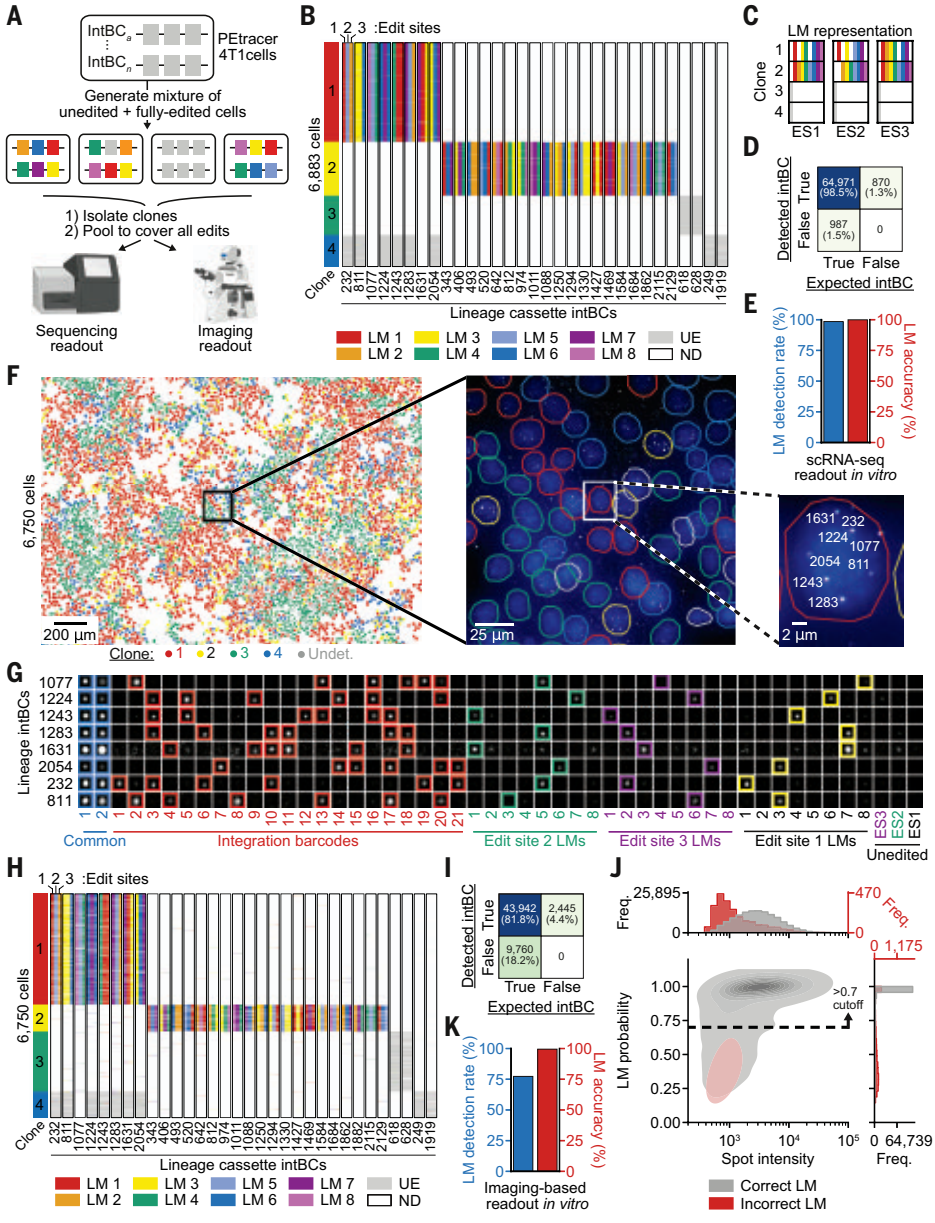


Fig. 2. Evaluation of PEtracer intBC and LM detection efficiency and decoding accuracy. (**A**) Experimental schematic for generating fully edited PEtracer cells with defined linkages between intBCs and LMs to enable ground-truth evaluation of intBC and LM detection efficiency and decoding accuracy. These clones have distinct sets of integrations, typically ~10 copies and thus ~30 ESs per clone. Representative sorting gates are shown in fig. S16. Sequencer and microscope images are from BioRender.com. (**B**) Droplet-based scRNA-seq of four clones (6883 cells). Color blocks on the left indicate clone identity: clone 1, red; clone 2, yellow; clone 3, green; clone 4, blue. LTCs consisting of three ESs are annotated with their respective intBCs at the bottom and show distinct LM combinations across the different clones. Unedited state (UE) is shown in gray; when an intBC was not detected (ND) in a cell, it is shown as white. (**C**) Summarized LM representation across the four combined clones. All 24 LMs are observed in the fully edited clones 1 and 2; the unedited state is represented in clones 3 and 4. (**D**) Confusion matrix depicting true positive, false negative, and false positive rates for detected and expected intBCs by droplet-based scRNA-seq. (**E**) LM detection rate (blue, left) and accuracy (red, right) for scRNA-seq data, based on expected intBC-LM pairing. (**F**) Widefield view (left) of fully edited and unedited clones plated for imaging-based lineage readout experiments, pseudocolored to indicate clonal identity. Color scheme matches that shown in (B), with the addition of undetermined cells (Undet.) that could not be confidently assigned to a clone shown in gray. Also shown are higher magnification of a field of view (middle) and a representative single cell (right) with 4′,6-diamidino-2-phenylindole (DAPI) nuclear staining (blue, integration amplicons as colocalized common bit puncta (white), and decoded intBCs numbered to match our codebook (see table S6). (**G**) Example readouts for intBC and LM detection and decoding after in situ T7 transcription of integrated LTCs. Seventeen imaging rounds in three colors (50 total bits, columns) for distinct integration amplicons (rows) include two common bits to identify candidate spots followed by 21 integration barcode bits (Hamming distance 4, Hamming weight 6 code), 24 LM bits, and three bits for the unedited state. (**H**) Same as shown in (B) but for imaging-based readout of 6750 fully edited cells using in situ T7 transcription of LTCs. (**I**) Confusion matrix as shown in (D) but for imaging-based readout. (**J**) Validation and performance of the logistic regression classifier for LM assignment from imaging-based PEtracer data. Freq., frequency. Assignment probability cutoff of $P > 0.7$ for decoded LMs is noted as a dashed line. (**K**) LM detection rate and accuracy after filtering and decoding with the LM classifier as shown in (J).

ESs over time, demonstrating uniform editing rates throughout the duration of the experiment (table S10). Across both cell lines, we observed expected patterns of protospacer mismatches leading to reduced editing rates, with PAM-proximal mismatches showing the largest reductions (fig. S4D). Notably, the relative editing rates for each mismatch were consistent across both lines; however, the median edit rate was 1.74-fold faster in B16-F10 cells compared with 4T1 cells (fig. S4, E and F). We selected protospacer variants that tuned the desired editing rate to match the timescales for subsequent in vitro and in vivo experiments over 2 to 3 weeks (0.05 to 0.1 edits per ES per day) and 4 to 6 weeks (0.02 to 0.04 edits per ES per day), respectively (Fig. 1G). Notably, the protospacer variants screened here could support tracing across diverse biological systems and timescales, ranging from a single cell division even in rapidly dividing cells to many months, although cell type–specific differences in PE performance necessitate empirical selection of optimal variants.

In vitro validation of PEtracer readout using fully edited cells

To measure PEtracer intBC and LM detection efficiency and decoding accuracy using both sequencing and imaging readouts, we generated 4T1 cells with predefined intBC-LM linkages (Fig. 2, A to C, and table S11). We edited cells to completion and pooled four clones to represent all LMs and unedited states for each ES ("fully edited cells," Fig. 2, A to C). Across 6883 cells profiled by scRNA-seq, we observed a 98.5% true positive rate for intBC detection and correctly called LMs for detected integrations with 99.9% accuracy (Fig. 2, D and E, and fig. S5A).

To evaluate intBC and LM decoding by MERFISH-based imaging, we mapped 30-nt sequencing to 183-nt imaging barcodes for each integration using scRNA-seq (fig. S5B and table S11). Fully edited cells were plated at high density onto coverslips, and we generated localized RNA amplicons through in situ T7 transcription (*56*) of genomically integrated LTCs using a modified in-gel transcription protocol, which integrates tissue clearing to enhance the accessibility of genomic loci and detection efficiency of RNA species (*92*, *93*) (Materials and methods). We then conducted 17 rounds of three-color imaging using hybridization probes to identify candidate amplicons; decode intBCs with a Hamming distance 4, Hamming weight 6 error-correcting MERFISH code; and detect the LMs and the unedited state for each ES (Fig. 2, F to H). Notably, this experiment allowed us to test all imaging readout bits. Across 6750 cells, our optimized protocol achieved an intBC detection rate of 81.8%, maintaining a low false positive rate and improving decoding performance for dimmer spots relative to published protocols (Fig. 2I; fig. S5, C to E; and tables S12 to S16). We next trained a logistic regression classifier to predict the most likely LM from spot intensities across imaging rounds, achieving 99.4% accuracy after excluding low-confidence LM calls, with >98.3% accuracy per LM (Fig. 2, J and K, and fig. S5, F to H). Our simulations indicate that this LM decoding performance supports accurate tree reconstruction.

Benchmarking of in vitro tree reconstruction

Next, we rigorously evaluated the accuracy of phylogenies reconstructed with PEtracer using scRNA-seq. To introduce independent ground-truth marks of shared ancestry, we performed two rounds of lentiviral static barcoding at sequential time points (day 5 and 7 of tracing) in 4T1 cells with ~10 to 15 integrated LTCs and a pegArray tuned to edit over a 2- to 3-week timescale (Fig. 3A). We then harvested cells for scRNA-seq, capturing evolving LM data, transcriptional profiles, and static barcodes for each cell.

We reconstructed phylogenies for six clonal groups of cells using only the evolving LM information (Fig. 3B and figs. S6 and S7A). These phylogenies show efficient LM labeling of each ES (ES1: 73.4%, ES2: 47.7%, ES3: 86.5%), with representation from all LMs (Fig. 3C). Notably, edit saturation is consistent across cells and independent of PEmax RNA levels, indicating stable expression (fig. S7B), and unintended edits occur at extremely low rates ("other"; ranging from 0.07 to 0.54%), demonstrating high editing fidelity (Fig. 3, B and C). To assess the quality of reconstructed trees, we quantified the agreement between groups of cells with a shared static barcode and phylogenetic clades defined by the inferred lowest common ancestor (LCA) of each group (Materials and methods). We reproducibly observed near-perfect agreement between LCA clades and static barcoding groups quantified with the Fowlkes-Mallows index (FMI) (*94*) ($0.85 \leq$ FMI scores ≤ 1.00) (Fig. 3, D and E; figs. S6 and S7A; and table S17). Both day 5 and 7 static barcode groups have consistently high FMI scores, demonstrating high accuracy at multiple depths within reconstructed phylogenies (fig. S7A).

Given the consistent editing rates demonstrated in our kinetics experiments (Fig. 1G), we could estimate the life span of each ancestral node in the tree and use these branch lengths to infer when static barcoding events occurred in the experiment. Although these branch length estimates tend to be noisy, the inferred mean time for static barcode LCAs aligned with the timeline of the experiment for most clones (Fig. 3, D and F, and fig. S7A). Phylogenetic distances based on branch length estimates significantly outperformed tree-independent LM distances in separating static barcode groups and were therefore used for all subsequent analyses (fig. S7, C and D).

We performed down-sampling analyses to determine reconstruction accuracy as biologically relevant conditions vary. Across all experimentally generated phylogenies benchmarked by static barcoding, we found that reconstruction accuracy remained high, with >20 ESs and with a detection rate of >60%—cutoffs that agree with our simulation work and that we use for subsequent reconstructions (Fig. 3, G and H, and tables S18 and S19).

Imaging-based spatial lineage analysis of motile cells

To evaluate the quality and scalability of phylogenetic reconstructions generated by PEtracer with imaging-based readouts, we profiled the growth dynamics of highly motile 4T1 cells in vitro (*95*). We sparsely seeded cells with active lineage tracing components onto coverslips and allowed these cells to grow into colonies over 6 days (Fig. 3I and fig. S8A). From a single coverslip, we captured lineage information for 18,675 cells across 64 evolving clones, which had an average of 39.5 ± 10.1 ESs per clone (mean ± standard deviation), a mean edit fraction of $75.8 \pm 22.2\%$ per clone, an intBC detection rate of $76.4 \pm 16.7\%$ per cell, and up to 1167 cells (Fig. 3, I and J; fig. S8, A to D; and table S20). Colonies can be seeded by individual cells or by multiple cells that grow into one another. Indeed, in these colonies, we observed considerable intermixing between seeding clones within polyclonal colonies, highlighting substantial cell movement during colony expansion (Fig. 3I). Similarly, we observed marked intermixing between distinct phylogenetic clades within clones (Fig. 3, K and L). In one representative clone, we observed that early LMs are widely distributed, whereas LMs that occur later are restricted to progressively smaller spatial regions, reflecting the limited time for dispersion because these cells shared a more recent common ancestor (Fig. 3M). Indeed, phylogenetic and spatial distances are correlated, with mean spatial distances between pairs of cells that share a common ancestor within the past 3 days being substantially lower than those between randomly permuted pairs (Fig. 3N and fig. S8, E and F). This trend is consistent irrespective of detection rate, demonstrating the accuracy of the phylogeny, a conclusion further supported by the correlation between phylogenetic and LM distances (a tree-independent metric) as well as our simulation work (fig. S1 and fig. S8, F and G). Direct observation of these colonies over time substantiated clones intermixing and rapid cell movement during outgrowth (movie S1). These results indicate that PEtracer imaging data can be used to reconstruct large, highly accurate phylogenies with true single-cell spatial resolution.

Spatially resolved transcriptomic profiling of metastatic tumors in vivo

We next applied PEtracer in vivo to jointly profile cellular state, microenvironment, and phylogeny using the syngeneic 4T1 breast carcinoma

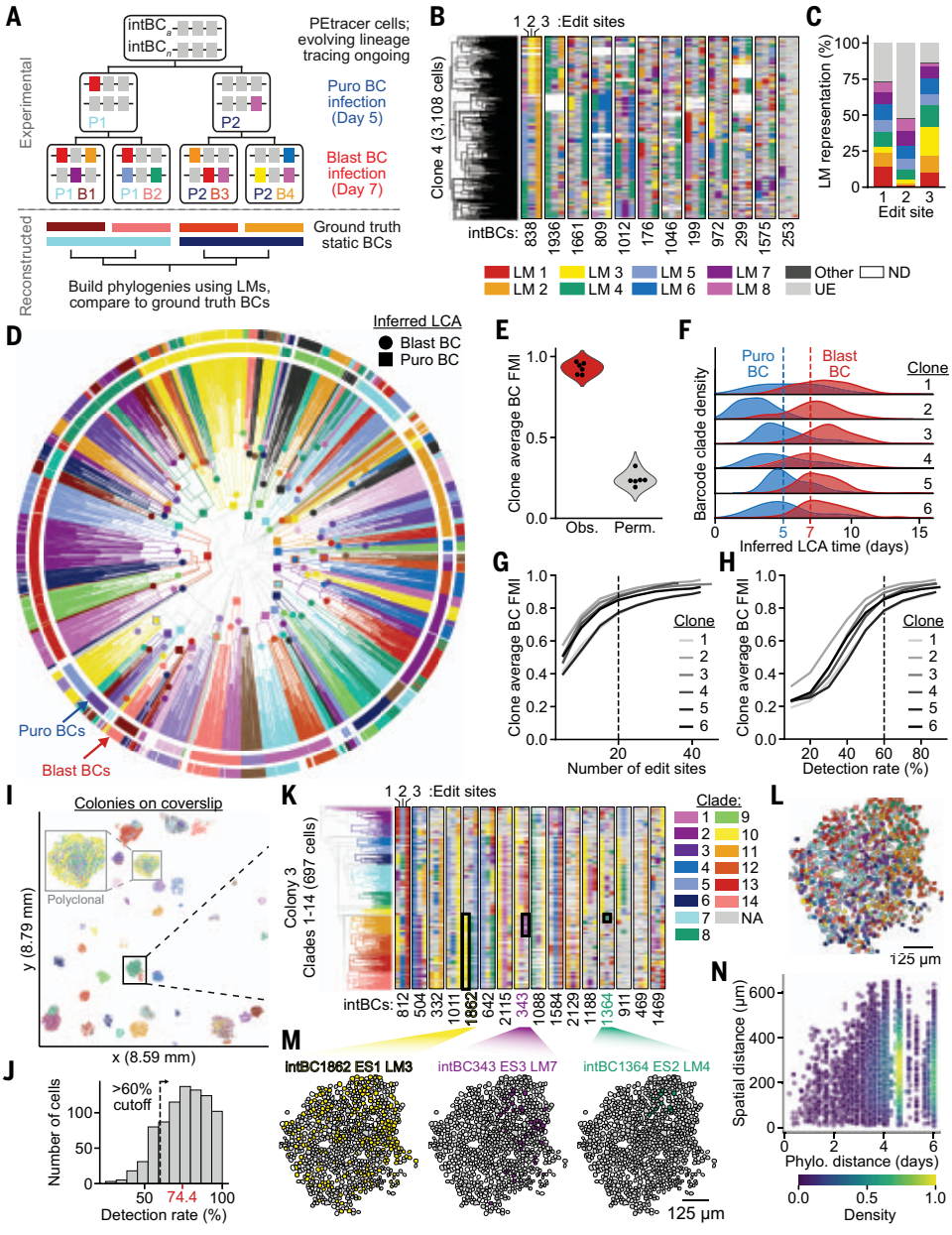


Fig. 3. Reconstruction of high-resolution, accurate phylogenies with PEtracer. (**A**) Schematic of an in vitro benchmarking experiment in which phylogenies are reconstructed based on the continuous accrual of LMs and reconstruction accuracy can be evaluated based on correct grouping of cells that share static barcodes introduced at two time points. (Top) Experimental design, in which cells engineered with PEtracer components begin evolving lineage tracing on day zero, followed by two rounds of lentiviral infection with static barcode (BC) libraries associated with antibiotic selection markers puromycin (Puro; blue lettering) and blasticidin (Blast; red lettering). (Bottom) Depicts how ground-truth static BC groups can be compared with reconstructed phylogenies. (**B**) Character matrix captured by scRNA-seq readout paired with phylogenetic tree reconstructed from LMs, using the neighbor-joining algorithm for the 3108 cells in clone 4. Each intBC is listed below its set of three corresponding ESs that are colored by LM identity. UE, unedited; ND, not detected. (**C**) Averaged ES modification frequencies for six clones assayed in this experiment, colored according to LM. (**D**) Phylogeny from (B) depicted as a circle with rings colored to show puromycin (inner ring) and blasticidin (outer ring) static BC groups for each cell (white indicates missing data). Phylogeny branches are colored to match BC assignments. The inferred lowest common ancestor (LCA) is marked and colored to match each BC group (Puro, square; Blast, circle). (**E**) Averaged Fowlkes-Mallows Index (FMI) scores for puromycin and blasticidin BC groups, measuring the agreement between LCA clades and observed (Obs.) static barcode labels versus randomly permuted (Perm.) barcode labels across the six phylogenies captured in this experiment. (**F**) Inferred LCA timing for puromycin and blasticidin static BC groups across six phylogenies. Dashed Day 5 and Day 7 lines indicate the true timing of when these BCs were introduced into cells. Averaged puromycin and blasticidin BC FMI scores are shown when down-sampling the number of ESs (**G**) or the detection rate (**H**) used for tree reconstruction with six experimental phylogenies assayed using scRNA-seq. (**I**) Imaged coverslip of 4T1 cells growing as intermixed clones with cells pseudocolored to indicate clone identity. *x* and *y* dimensions are provided in millimeters. The gray inset box highlights a polyclonal colony comprising clones 2, 14, 15, and 29. The black inset box shows clone 3 of 64, which we analyze in (J) to (N). (**J**) Histogram of intBC detection rate for cells in clone 3. The average detection rate is shown in red; a 60% cutoff for cells included in phylogenetic reconstructions is noted with a dashed line. (**K**) Character matrix captured by imaging readout of the 697 cells, with robust intBC and LM detection in clone 3. LM annotations identical to those shown in (B) are provided across the intBCs enumerated on the *x* axis. Clades 1 through 14 are colored on the phylogeny to the left. NA, not assigned a clade. (**L**) Nuclear masks for cells in clone 3 colored according to their phylogenetic clade assignment. (**M**) Spatial positions of cells in clone 3 marked with specified edits that define spatial neighborhoods of cells. (**N**) Pairwise phylogenetic (Phylo.) versus spatial distance for cells in clone 3.

cell line, which recapitulates cancer cell extravasation, metastatic colonization, and growth in immunocompetent mice (*89*, *90*, *96*). We engineered a clonally diverse population of 4T1 cells with ~10 to 15 integrated LTCs and a pegArray tuned to edit over 4 to 6 weeks. scRNA-seq of these cells during in vitro culture revealed two distinct transcriptional states, distinguished by differential expression of epithelial-to-mesenchymal transition (EMT)–related genes (fig. S9, A and B). Clonal lung metastases were then initiated by tail vein injection of tracing cells sorted by FACS either 6 or 12 days after lentiviral induction of PEmax expression (Fig. 4A).

After tumor outgrowth in vivo, we harvested the lungs of tumor-bearing mice and collected 20-μm tissue sections. Extensive experimental optimizations to prevent RNA degradation and accurately segment cells (*97*–*102*) enabled MERFISH imaging with a 124-gene library designed to annotate major cancer, immune, and stromal cell types (*103*–*105*) (Fig. 4A, table S21, and Materials and methods). Across five sections from two mice, we captured 10 distinct tumors, all of which were derived from a more mesenchymal clone observed in vitro (fig. S9, C to E). In total, we profiled 368,722 cells (mean: 178 transcripts per cell) with high correlation between MERFISH replicates and with bulk and scRNA-seq data (Fig. 4, B to D; fig. S10, A to D; and table S22) (*106*). Eighteen cell types were annotated on the basis of the expression of previously described marker genes (Fig. 4, D and E, and fig. S10, E and F) (*103*–*105*). Overall, we observed consistent cell type representation across tumors but variable frequencies related to tumor size, with less immune infiltration in larger tumors (Fig. 4D). Larger tumors often had necrotic cores containing cellular debris, with few intact nuclei or RNA molecules detected by MERFISH (Fig. 4C and fig. S10A).

We next defined five cell type neighborhoods within 4T1 tumors using Hotspot (*107*) (Fig. 4F and fig. S11) and examined cell type densities relative to the lung-tumor interface, which we found to be the most informative axis of spatial organization (Fig. 4G). A lung-endothelial neighborhood with increased AT1/AT2, club, and endothelial cell densities were predominantly localized in nontumor, normal lung tissue, whereas the tumor core mainly consisted of malignant cells (Fig. 4, F to H). We identified an immune-enriched neighborhood located near the lung-tumor interface, although we observed immune infiltration throughout the tumor (Fig. 4, F to H). We noted focal B cell–enriched

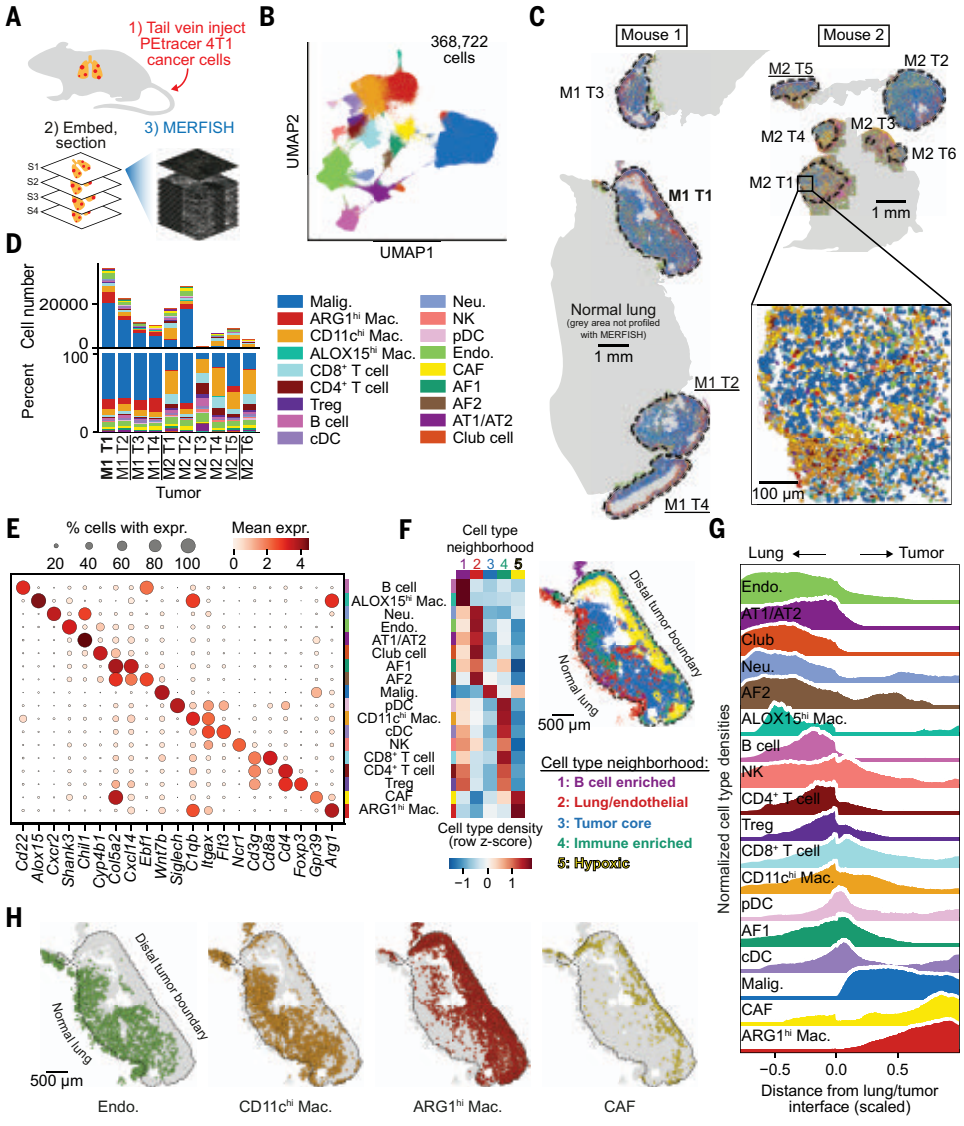


Fig. 4. Spatial transcriptomics of metastatic 4T1 lung tumors. (**A**) Experimental schematic for generating MERFISH data from lung tumors after tail vein injection of 4T1 cells engineered with PEtracer components into immunocompetent mice. PEtracer 4T1 cells undergo evolving lineage tracing throughout tumor growth; tumor-bearing lungs are embedded for sectioning and imaged. (**B**) Uniform manifold approximation and projection (UMAP) embedding of 368,722 cells sampled from 10 tumors in two animals, where each cell is colored according to its assignment to the 18 distinct cell types resolved by this library. Cell type color annotations are listed below the UMAP. (**C**) Spatial organization of cell types in 4T1 lesions in the lungs of two mice, with a zoom-in showing nuclear segmentation masks. Cell types are colored with the annotations shown in (B). Mouse 1 (M1; left) has four distinct tumors (T1 to T4); Mouse 2 (M2; right) has six (T1 to T6). M1 T1 is bolded to indicate that this tumor is analyzed more deeply in Fig. 5 and fig. S14. M1 T2, M1 T4, and M2 T5 are underlined because they are analyzed further in fig. S13. Lung boundary is show in gray, and normal lung cells profiled by MERFISH (on average 154 μm from tumor boundary, indicated by dashed line) are colored according to cell type annotation; tissue-level scale bars, 1 mm; insert scale bar, 100 μm). (**D**) Percentage and cell number counts for each cell type grouped by tumor; colors are consistent with those shown in (B). (**E**) Marker gene expression (expr.) across annotated cell types. Dot color indicates the mean gene expression for a marker gene within a cell type, and dot size indicates the percentage of cells within a cell type that express that gene. (**F**) Heatmap of cell type densities in neighborhoods identified by Hotspot (left, row *z*-scored). Spatial distribution of cell type neighborhoods for M1 T1 section 2 (right). (**G**) Normalized cell type densities as a function of their scaled distance to the lung-tumor interface (boundary between adjacent lung and tumor; −1 corresponds to cells within the lung furthest from the lung-tumor interface, and +1 corresponds to cells within the tumor furthest from the lung-tumor interface). Cell type colors are consistent with legend in (B). (F) and (G) depict data from nine tumors (M2 T3 was excluded from analysis because of low cell number). (**H**) Spatial distribution of endothelial cells, CD11c^{hi} macs., ARG1^{hi} macs., and CAFs across M1 T1 section 2. Malig., malignant; Mac., macrophage; Treg, regulatory T cell; cDC, conventional dendritic cell; Neu., neutrophil; NK, natural killer cell; pDC, plasmacytoid dendritic cell; Endo., endothelial cell; CAF, cancer associated fibroblast; AF1/AF2, alveolar fibroblast type 1/2; AT1/AT2, alveolar epithelial type 1/2.

www.xsava.xyz

neighborhoods located outside the tumor boundary that were also enriched for arachidonate 15-lipoxygenase-high (ALOX15^{hi}) macrophages (Fig. 4, F to H, and fig. S11). Arginase 1-high (ARG1^{hi}) macrophages and cancer-associated fibroblasts (CAFs) were enriched at the tumor boundary furthest from the adjacent lung (distal tumor boundary; Fig. 4, G and H) (*108*). We also observed reduced endothelial cell density within this hypoxic neighborhood at the distal tumor boundary, which is consistent with prior descriptions of M2-like macrophages enriched in hypoxic tumor regions (*109*–*112*) (Fig. 4, F to H). These results reveal the intricate spatial organization of diverse cell types within the tumor microenvironment in this syngeneic model of cancer metastasis.

Lineage maps reveal spatial patterns of metastatic tumor growth

To dissect how cell lineage informs tissue structure, we performed our optimized in-gel T7 amplification protocol to read out lineage information in MERFISH-processed tumor sections (Fig. 5A). We first evaluated intBC and LM decoding in vivo by imaging in a primary 4T1 tumor comprising fully edited cells, where we observed 99.6% LM readout accuracy in tumor sections (fig. S12). We noted a relationship between intBC detection efficiency, nuclei volume, and nuclei *z* position (Fig. 5B). Decreased nuclei volumes at the top and bottom of the tissue slice suggest bisection of nuclei during sectioning; we therefore only included nuclei with centroid *z* positions >4 μm (corresponding to the average nuclei diameter) from the top or bottom of tissue sections—the average intBC detection rate for these cells was 55.2 ± 9.7% (fig. S13A and table S23). Four clonal tumors with mean edit fractions >50% and 36 to 39 ESs with a common set of intBCs were selected for further analysis (fig. S13, B and C, and fig. S9, C to E). Although the presence of tumors with low edit fractions suggests that PE silencing can occur at the clone level, potentially exacerbated by selection against immunogenic PEtracer components, the narrow distribution of edit fractions within clones indicates that subclonal silencing is rare (fig. S13B). As with the in vitro 4T1 colonies, we reconstructed phylogenies using all cells with >60% intBCs detected, on average 45.2% of whole malignant cell nuclei (range 35.8 to 57.5%; Fig. 5C and fig. S13D). After quality filtering, in vivo trees ranged in size from 1265 [Mouse 2 Tumor 5 Section 1 (M2 T2 S1)] to 19,029 cells (M1 T1 S1 to S4) with near-optimal editing kinetics and LM diversity, which is consistent with in vitro editing with these pegRNAs (Fig. 5, D and E, and fig. S13, E and F).

With these tissue-scale phylogenies, we examined the coordinated effects of cell state, lineage, and spatial organization on tumor evolution in vivo. Phylogenetic clades occupy distinct spatial regions with consistent patterns observed across multiple sections (Fig. 5, E and

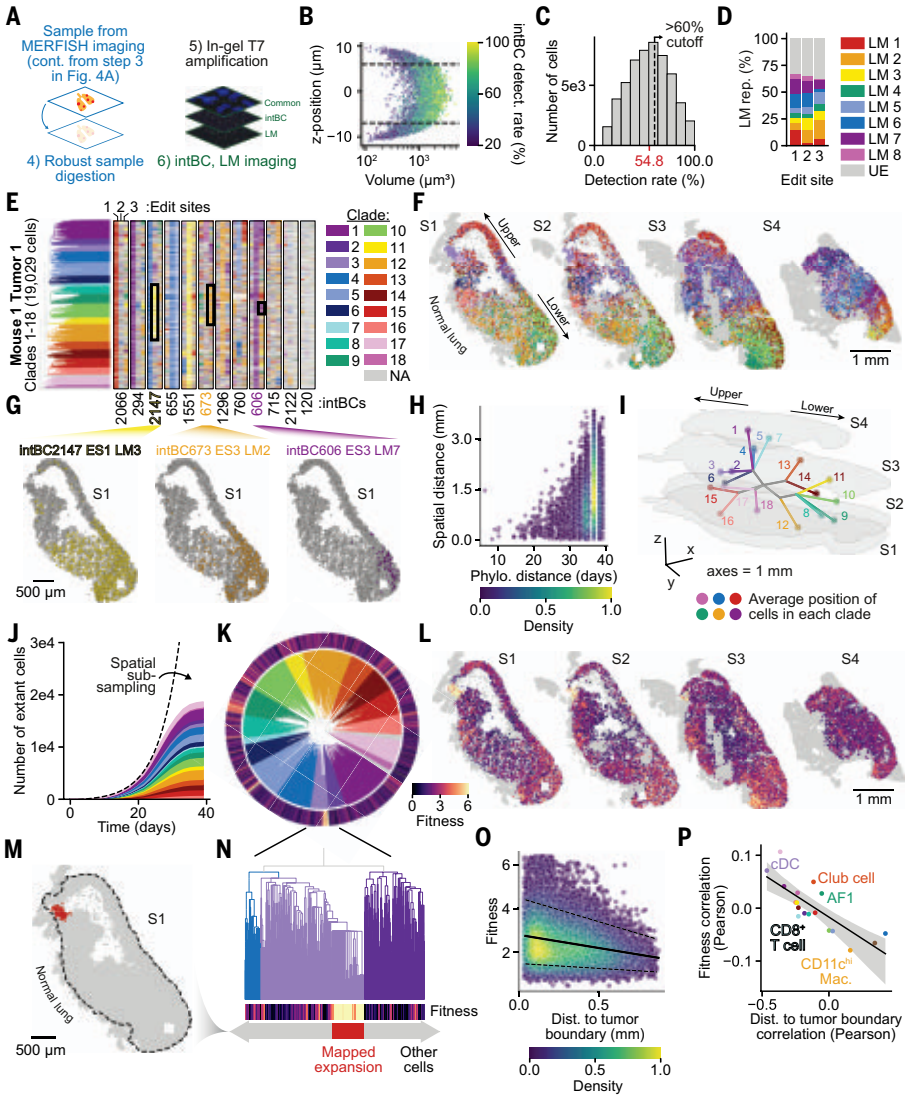


Fig. 5. In vivo phylogenetic reconstructions at tissue scale. (**A**) Experimental schematic for imaging-based readout of PEtracer lineage data from samples after completion of MERFISH RNA readout. Samples undergo 4) extensive digestion, 5) in-gel T7 in situ amplification, and 6) imaging of PEtracer intBCs and LMs. (**B**) intBC detection (detect.) rate as a function of cell volume and its *z*-position. (**C**) Histogram of cell count by in vivo intBC detection rate for M1 T1. The average detection rate is shown in red; a 60% cutoff for cells included for phylogenetic reconstructions is noted as a dashed line. (**D**) ES LM modification representation (rep.) across all cells included in the phylogeny for M1 T1. UE, unedited. (**E**) Character matrix and phylogeny for the 19,029 cells with >60% intBC detection efficiency across four sections (S1 to S4) for M1 T1. intBCs are listed below the character matrix, which is colored according to LM identity for each ES. LM coloring is same as shown in (D); additionally in (E), white indicates not detected. Clades 1 through 18 are colored in the phylogeny. NA, not assigned. Clade color scheme is shared with (F). (**F**) Cells from four sections (S1 to S4) colored by clade assignment. The location of the normal lung and the upper and lower portions of the lesion are annotated. Scale bar, 1 mm. (**G**) Spatial organization of edits in M1 T1 S1 that define phylogenetically related spatial domains. Gray cells denote other LMs or unedited; white cells denote not detected. (**H**) Pairwise phylogenetic (Phylo.) versus spatial distance for M1 T1 phylogeny. (**I**) Three-dimensional reconstruction of the colored clades in M1 T1. Balls indicate the average spatial position of cells in a clade; branches indicate the average position of contributing ancestors to each node. Sections S1 to S4 are depicted as light outlines. Axes length, 1 mm. (**J**) Estimated number of extant cells as a function of time, based on inferred branch lengths. Each clade is colored as shown in (E); the dotted line shows an exponential fit curve of tumor growth. (**K**) Phylogeny from (E) for M1 T1 cells across all four sections depicted as a circle, with phylogeny branches colored according to clade assignment. The outer ring depicts inferred fitness for each cell. (**L**) Spatial organization of cell fitness across four sections for M1 T1. A spatially confined (**M**), high-fitness expansion (**N**) is observed at the lung-tumor boundary. Cells in this expansion in S1 are shown in red; other cells are shown in gray. (**O**) Malignant cell distance (Dist.) to the tumor boundary versus fitness. The black line shows regression; dotted lines mark top and bottom deciles. (**P**) Pearson correlation between indicated cell type density with fitness versus correlation of cell type density with distance to tumor boundary. The black line depicts regression with ribbon indicating 95% confidence interval. Mac., macrophage, AF1, Alveolar fibroblast 1; cDC, conventional dendritic cell. See table S24 for relevant statistics.

F, and fig. S13G). In M1 T1, for example, intBC2147 ES1 LM3 marked an early split in the tree, with cells bearing this LM predominantly located in the lower half of the tumor, suggesting an early axis of tumor growth parallel to the lung-tumor interface (Fig. 5G). LMs defining later branches in the tree, representing cells with more recent common ancestors, are confined to progressively smaller spatial regions (Fig. 5G). Notably, although spatial intermixing of clades indicates that 4T1 cells remain motile in vivo, the correlation between pairwise phylogenetic distance and maximum spatial distance was higher than that in in vitro colonies, possibly reflecting reduced local cell movement due to increased cell densities in vivo (Fig. 5, F and H, and fig. S13G). Cells with more-similar sets of LMs (lower LM distances) also tended to be spatially close, although this trend was weaker than that observed for phylogenetic distance (fig. S14, A and B).

To perform a 3D reconstruction of tumor growth, we calculated the average spatial location of cells within each phylogenetic clade and that of ancestral nodes in the tree across multiple tissue sections of the same tumor (Fig. 5I). This analysis nominated an early phylogenetic and spatial bifurcation of tumor growth, with clades 1 to 7 located in the upper half of the tumor and clades 8 to 12 in the lower half. Clades 17 and 18 represent a phylogenetic outgroup with rich lineage substructure that is spatially enriched at the lung-tumor interface in the upper region of the tumor, a region with high local phylogenetic diversity (Fig. 5, F and I, and fig. S14, C to E). Estimations of extant cell number for each clade at various time points suggested exponential growth for all clades throughout tumor progression, which is consistent with mostly neutral evolution at the clade level (Fig. 5J). We expect that divergence from exponential expansion at later time points is caused by subsampling due to the limited number of tumor sections that we collected, despite high cell recovery for individual sections, although subsampling could also mask changes in growth rates later in tumor development.

To understand drivers of cell lineage dynamics in vivo, we used tree structure to assess phylogenetic fitness, an integrated measure over time of either increased cell division or decreased cell death relative to other subclades. We calculated fitness for each cell (Materials and methods) and examined the spatial distribution of high-fitness cells, which we found to be enriched along the tumor periphery adjacent to the normal lung, including a highly localized expansion at the lung-tumor interface (Fig. 5, K to N) (*54*). Overall, we observed a negative correlation between phylogenetic fitness and distance to the tumor boundary (Fig. 5O), a trend also observed by using mean neighbor LM distance (a tree-independent fitness metric) (fig. S14, F to H). Given the diverse spatial distribution of immune and stromal cells, we asked whether local cell type densities correlated with cancer cell fitness. Although the density of several cell types [e.g., conventional dendritic cells (cDCs)] were weakly correlated with fitness, we found that the densities of fitness-associated cell types also varied by their distance to the tumor boundary, complicating efforts to disentangle cell type density from spatial position effects (Fig. 5P; fig. S14, I to K; and table S24). The wide variability in cancer cell phylogenetic fitness that we observed at the tumor boundary suggests that additional cell-intrinsic and -extrinsic factors may cooperate to influence cancer cell dynamics and fitness.

Transcriptional and spatial determinants of cancer cell fitness

To better characterize transcriptional and spatial features that underlie differences in cancer cell fitness, we augmented our MERFISH library to assess gene programs associated with cancer progression (e.g., hypoxia, EMT, and cell cycle). Using this expanded 175-gene library, we generated MERFISH data from three sections of an additional tumor, assaying 104,219 total cells (mean: 261 transcripts per cell) (M3 T1; tables S22 and S25). Our new library recovered previous annotations and further resolved previously detected cell types into transcriptionally distinct subtypes (fig. S15, A to D). In addition, we identified four spatially coherent transcriptional modules of malignant cells using Hotspot (*107*): (i) "leading edge" cells at the distal tumor boundary, (ii) "hypoxic" cells, (iii) "tumor core" cells, and (iv) "lung adjacent" cells (Fig. 6, A to C, and fig. S15E). These modules were also associated with distinct cellular microenvironmental neighborhoods (fig. S15F).

To understand how cancer cell modules emerged during tumor growth, we constructed a phylogeny for the 10,061 cells with >60% intBC detection in M3 T1 (Fig. 6D and fig. S15G). Each cancer cell module had mostly uniform clade compositions, except the lung adjacent module which was enriched for clades 13 to 16 (Fig. 6, E and F). We observed intermixing of cancer cell module assignments within phylogenetic clades, suggesting that 4T1 cells exhibit transcriptional plasticity to adapt to microenvironmental conditions (Fig. 6D). Despite this plasticity, we find striking differences in cancer cell module heritability. Specifically, the lung adjacent module was the most heritable (Fig. 6G and table S26), which is possibly related to high expression of epithelial markers (e.g., *Cdh1*; E-cadherin) in contrast to leading edge cells, which up-regulate mesenchymal markers (e.g., *Snai1*; snail family zinc finger 1) and therefore are more likely to move and exhibit transcriptional plasticity (Fig. 6B) (*113*). We next used the M3 T1 phylogenetic tree to parse how module-associated features coordinately instruct cancer cell fitness. As before, high-fitness cells were overrepresented in certain phylogenetic clades and enriched at the tumor boundary, particularly within the lung adjacent module (Fig. 6, D, H, and I). Notably, the fraction of ES edited per cell shows no clear correlation with either cancer cell fitness or module (fig. S15, H and I, and table S23).

These data suggest that cancer cell modules reflect the integrated effects of transcriptional programs and environmental cues that instruct tumor growth. Leading edge cells show higher expression of cell cycle–related genes [including *Foxm1* (forkhead box M1), a transcription factor previously identified as a genetic dependency for 4T1 growth in vivo] (*114*) and occupy a cellular neighborhood enriched for CAFs at the distal boundary (Fig. 6, J to L; Fig. 4, G and H; and fig. S15F). The spatial and transcriptional signature of leading edge cells has low heritability, indicating that the proliferative capacity of these cells is a transient state influenced by local factors that do not instruct long-term behaviors (Fig. 6, E and G). A hypoxic microenvironment beneath leading edge cells defines module 2, marked by expression of angiogenesis-related genes (e.g., *Vegfa;* vascular endothelial growth factor A) and a sharp boundary between tumor endothelial cells and $ARG1^{hi}$ macrophages (Fig. 6, M and N) (*109*–*112*). Tumor core malignant cells show no distinct transcriptional or lineage signatures. By contrast, lung adjacent cells have the highest fitness, and this module is the most heritable in this tumor (Fig. 6, G and I). Phylogenetic fitness reflects proliferative potential across several weeks and many cell divisions, whereas cell cycle state can fluctuate in a matter of hours, which we suspect is why lung adjacent cells have higher fitness than leading edge cells despite having a slightly lower proportion of actively cycling cells (Fig. 6L).

We next sought to identify the drivers of high-fitness cancer cell states, including lineage, positional, cell-intrinsic, and cell-extrinsic factors. Similarly to our observations in M1 and M2 tumors, we found that the densities of several cell types [e.g., regulatory T cells (Tregs)] were positively correlated with phylogenetic fitness and also enriched at the tumor boundary near lung adjacent cells (fig. S15, F and J). To identify cell-intrinsic drivers of cancer cell growth, we looked for fitness associated genes that also exhibited heritable expression patterns (Fig. 6O and tables S24 and S26). This analysis nominated genes associated with an epithelial-like state that are highly expressed in high-fitness lung adjacent malignant cells but lowly expressed elsewhere, including *Cldn4* (Claudin 4, an epithelial cell tight junction protein) and *Fgf1*/*Fgfbp1* (fibroblast growth factor 1 and its carrier protein FGF binding protein 1) (Fig. 6, O to R). To disentangle the contributions of spatial cues versus heritability in these cells, we quantified the extent to which gene expression patterns were explained by spatial proximity

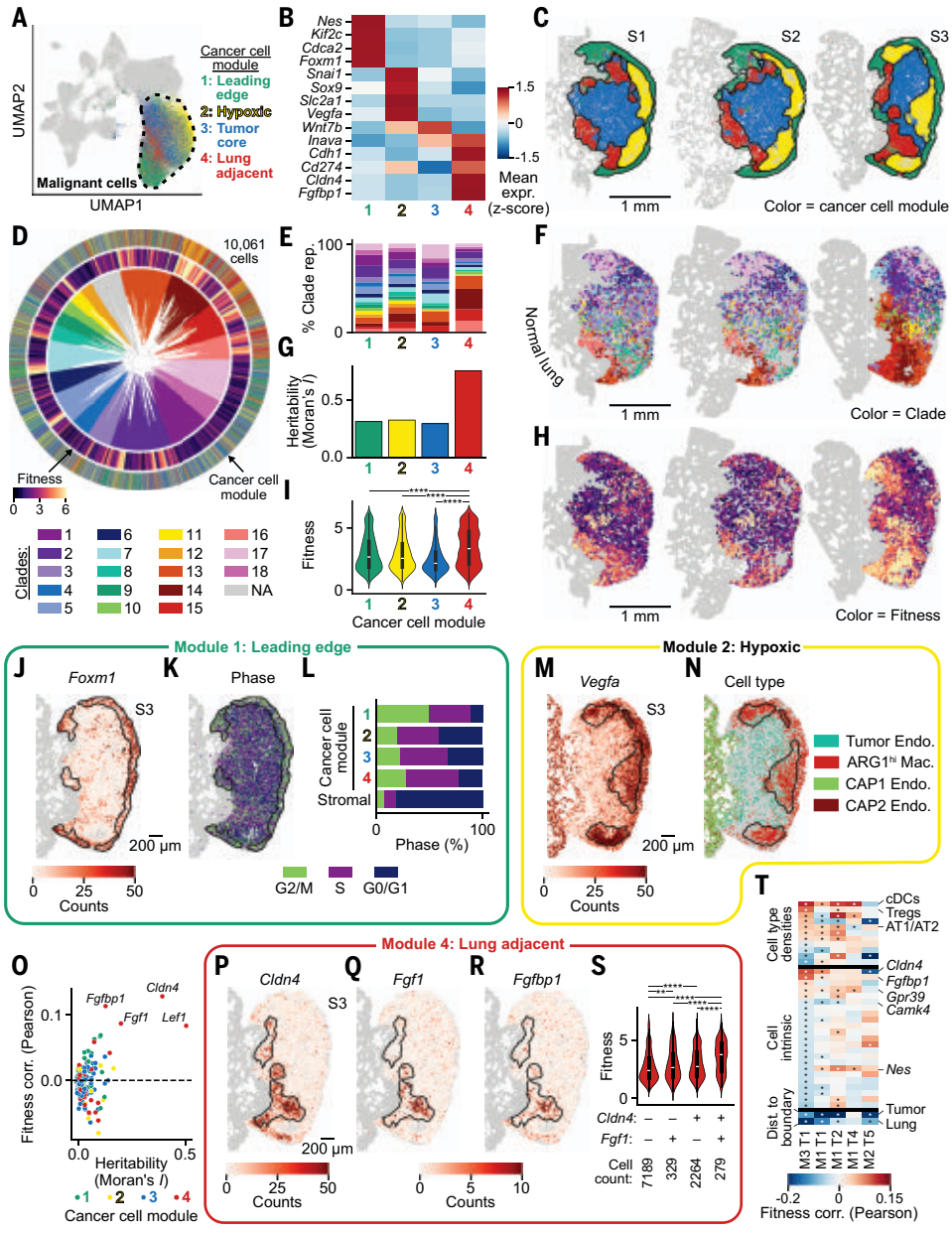


Fig. 6. High-resolution spatial mapping of cell state and lineage uncovers drivers of tumor growth. (**A**) UMAP embedding of 104,219 cells sampled across three sections (S1 to S3) of a tumor from Mouse 3 (M3 T1), in which the malignant cells are outlined and colored by their assignment to four distinct cancer cell modules: Module 1, leading edge; Module 2, hypoxic; Module 3, tumor core; and Module 4, lung adjacent. The full UMAP embedding, in which each cell is colored by its annotated cell type, can be found in fig. S15A. (**B**) Heatmap of marker gene expression (expr.) for each cancer cell module. (**C**) Spatial organization of modules 1 to 4 across M3 T1 S1 to S3. (**D**) Phylogeny for all malignant cells with >60% integration BC detection efficiency across M3 T1 S1 to S3, depicted as a circle. Branches are colored by clade assignment (Clades 1 to 18) with a color key beneath. The inner ring denotes fitness, and the outer ring denotes the module assignment for each of the cells in the tree. (**E**) Representation (rep.) of each of the 18 phylogenetic clades across the four cancer cell modules. (**F**) Spatial organization of phylogenetic clades 1 to 18 across M3 T1 S1 to S3. (**G**) Quantification of module assignment heritability as measured by Moran's *I* phylogenetic autocorrelation. (**H**) Spatial organization of cell fitness scores across M3 T1 S1 to S3. (**I**) Violin plots of cell fitness scores by module, with embedded box plot showing quartiles. (**J** to **L**) Module 1/leading edge cell analyses. (J) MERFISH counts for *Foxm1* transcripts in M3 T1 S3 (malignant cells only). (K) Spatial organization cell cycle phase in M3 T1 S3 (malignant cells only). (L) Quantitation of cell cycle phase by cancer cell module in contrast to stromal cells in the adjacent normal lung. (**M** and **N**) Module 2/hypoxic cell analyses. (M) MERFISH counts for *Vegfa* transcripts in M3 T1 S3 (all cells). (N) Spatial distribution of tumor endothelial (Endo.), ARG1^{hi} macrophages (Mac.), and capillary 1/2 (CAP1/CAP2) Endo. cells. (**O**) Heritability versus Pearson fitness correlation (corr.) of individual genes colored according to module assignment. (**P** to **S**) Module 4/lung adjacent cell analyses. MERFISH counts are shown for (P) *Cldn4*, (Q) *Fgf1*, and (R) *Fgfbp1* transcripts in M3 T1 S3. (S) Violin plot is shown for Module 4 malignant cell fitness as a function of *Cldn4* and *Fgf1* expression. Abundance of cells with varying states is listed below each entry. (**T**) Pearson correlation of malignant cell fitness with cell type densities, cell-intrinsic gene expression, or tumor features [distance (dist.) to the tumor boundary or lung-tumor interface] across the five tumors analyzed in this study. cDC, conventional dendritic cell, Treg, T regulatory cell, AT1/AT2, alveolar epithelial type 1/2. Scale bars, 1 mm for (C), (F), and (H), and 200 µm for (J) to (K), (M) to (N), and (P) to (R). Significance in (I) and (S) denotes Mann-Whitney-Wilcoxon two-sided test where $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$. Significance in (T) denotes Student's *t* test of Pearson correlation coefficient. See tables S24, S26, and S27 for relevant statistics.

versus phylogenetic relatedness. *Cldn4* expression was highly heritable, indicating a stable cell-intrinsic state, whereas *Fgf1*/*Fgfbp1* expression was more spatially restricted and less heritable, which is consistent with their signaling functions (Fig. 6, O to R, and fig. S15K). Notably, *Fgf1* is also expressed by nonmalignant cells (fig. S15C), suggesting both paracrine and autocrine signaling. These genes illustrate how spatially and phylogenetically coherent programs interact to promote lung adjacent cell fitness, where cells coexpressing both *Cldn4* and *Fgf1* exhibit the highest phylogenetic fitness (Fig. 6S). These findings are reproducible across tumors, with similar sets of cell-extrinsic and -intrinsic factors associated with fitness, including *Cldn4* and *Fgfbp1* expression (Fig. 6T, fig. S15L, and table S24).

Discussion

Here, we describe PEtracer, an evolving lineage tracing technology that enables joint phylogenetic and cell state profiling with either single-cell sequencing or high-resolution imaging while preserving cellular tissue context. Guided by computational modeling, we engineered PEtracer to enable robust, high-resolution phylogenetic reconstructions capable of marking individual cell divisions in fast-dividing cells and tuned editing kinetics to enable lineage recording over many weeks or months. PEtracer allows us to capture key events that occur dynamically throughout biological processes without a priori knowledge of event timing. We performed rigorous benchmarking of PEtracer to establish that we can assemble high-quality phylogenies that accurately reconstruct static barcode groups across multiple time points. Using MERFISH-based imaging—chosen for its high spatial (single-cell and subcellular) resolution, high RNA detection efficiency, and minimal cell loss relative to current sequencing-based spatial methods—we collected rich cell state and lineage data from intact tissue sections (*115*). This approach preserved the cellular tumor microenvironmental context and enabled high-resolution reconstruction of growth patterns in individually seeded tumors within a transplantable model of murine breast cancer lung metastasis (*89*, *90*, *96*). PEtracer is also well-suited for integration with other imaging- and sequencing-based spatial transcriptomic methods (*55*, *116*–*118*).

Our tumor metastasis studies highlight how integrating cell state, lineage, and spatial information at tissue scale in vivo provides unprecedented insights into tumor evolution. Using MERFISH, we identified spatially and transcriptionally coherent cancer cell modules defined by distinct cell-intrinsic and -extrinsic programs. At the outer tumor edge, increased malignant cell division is likely driven by space for further growth and a supportive cellular neighborhood rather than heritable factors because phylogenetic analysis reveals low heritability of this state. Notably, phylogenetic fitness and cell cycle state reflect proliferative capacity over different time scales, and leading edge cells do not have the highest fitness within the tumor despite a modestly higher proportion of proliferating cells. As leading edge cells divide, they create a hypoxic, high-density cell layer where low oxygen availability induces up-regulation of mesenchymal gene expression programs (*119*) and high *Vegfa* expression, which promotes angiogenesis (*120*). Within the more vascularized tumor core, spatial constraints limit growth despite an available blood supply. Conversely, epithelial-like malignant cells adjacent to the normal lung adapt to a nutrient- and immune-rich niche, upregulating heritable transcriptional programs (e.g., *Cldn4* and *Fgf1*/*Fgfbp1*) associated with high phylogenetic fitness. These adaptive programs differ from the more mesenchymal states that enable extravasation and initial seeding of lung lesions (*121*). Spatially resolved phylogenies reveal complex relationships between transcriptional state, spatial location, and lineage that would be difficult to resolve by using only spatial or transcriptional neighbors to make lineage inferences, owing to cell migration and phenotypic plasticity. Together, our results highlight the interplay between heritable transcriptional programs and the tumor microenvironment in shaping cancer cell fitness and evolutionary dynamics.

These efforts motivate several new directions to further understand evolutionary relationships between cells in complex biological systems: (i) Although our work here describes a single model of cancer cell extravasation and metastatic outgrowth, other tumor models better capture cancer initiation, progression, therapeutic resistance, and the complete metastatic cascade. As with our previous Cas9-based lineage tracing system—which shares similar core components—we envision using PEtracer to study a range of tumor models as well as development (*13*–*16*, *39*, *53*, *54*). (ii) Assaying tissues with both imaging and sequencing readouts (*116*–*118*) would enable full transcriptomic imputation using phylogenetic neighbors (*122*–*125*). (iii) Imaging of thicker tissue sections with concomitant computational improvements would better enable 3D tissue reconstruction (*126*). (iv) Novel lineage reconstruction algorithms may further improve phylogenetic reconstructions (*127*–*130*). (v) Layering perturbations on top of phylogenetic reconstructions would enable direct interrogation of the drivers of cell state transitions and evolutionary dynamics (*131*). (vi) Integrating emerging PE-based signal recording methodologies (*132*) would enable decoration of phylogenies with LMs associated with signaling events of interest to assess their timing and impact on fate determination.

More broadly, understanding how cellular histories shape tissue-scale characteristics is a long-standing challenge in biology. The PEtracer system enables direct interrogation of cellular dynamics across diverse biological contexts, complementing descriptive cell atlases (*8*, *133*–*142*) to understand the drivers of cell fate determination and tissue organization.

Materials and methods

In silico lineage tracing simulation

Lineage trees and associated tracing data were simulated using the Cassiopeia framework (v2.0.0) (*72*). A birth-death process was employed to generate lineage trees, where internal nodes represented cell divisions, branch lengths corresponded to individual cell lifetimes, and leaves represented the cells observed at the experiment's conclusion. The time between cell divisions was modeled using a log-normal distribution (mean = 1, sigma = 0.5) and the cell death rate was set at 0.25. Each simulation was run until the desired population size was reached or all cells died. These simulated trees were subsequently used to generate tracing data under varying conditions.

To evaluate how key experimental parameters influence the accuracy of reconstructed trees, tracing data was simulated while varying the following parameters:

- N: The number of extant cells at the end of the experiment
- M: The number of ESs available for recording information
- S: The number of unique LMs that could be installed at each ES
- X: The distribution of LM installation probabilities
- F: The fraction of ESs with an LM installed at the end of the experiment
- D: The efficiency of LM detection at each ES

Simulations began at the tree's root with M unedited sites. Edits occurred stochastically and were inherited according to the tree structure. To achieve the target edit fraction F over a total experimental time T, edits occurred at a rate R given by:

$$R = 1 - (1-F)^{\frac{1}{T}}$$

The LM installed during each edit was randomly selected based on the distribution X. At the end of the simulation, a fraction $1 - D$ of the LMs was randomly dropped out to account for imperfect detection.

Distance-based tree reconstruction

For tree reconstruction with neighbor joining and UPGMA a N-by-N weighted Hamming distance matrix was computed. The distance between two cells i and j is defined as:

$$d(i,j) = \frac{\sum_{m=1}^{M} \delta(C_{im}, C_{jm})}{L}$$

Here, C represents the N cell by M ES matrix of detected LMs, L is the number of ESs where both cells have an LM detected, and the distance for ES m is defined as:

$\delta(c_i, c_j) = 0$ if c_i matches c_j or either is missing

1 if c_i or c_j is unedited

2 otherwise (both edited and not equal)

To efficiently reconstruct large trees, optimized versions of neighbor joining and UPGMA were used (*78*). These algorithms employ dynamic programming to achieve an $O(n^2)$ runtime and are implemented in the CCPhylo software suite (v0.8.3, https://github.com/genomicepidemiology/ccphylo).

For rooting trees generated by neighbor joining, a synthetic cell with all sites unedited was added. The tree was then rooted using this synthetic cell as the outgroup, after which the synthetic cell was removed.

Greedy tree reconstruction

We also evaluated tree reconstruction using the "vanilla" greedy algorithm from Cassiopeia (*72*). This algorithm builds a tree from the top down by recursively selecting lineage marks (LMs) that split the cells into progressively smaller groups. LMs are chosen based on their frequency within the group, with lower-frequency LMs favored because they are less likely to exhibit homoplasy (i.e., multiple independent occurrences of the same edit at the same ES in separate cells). Although the greedy algorithm performed well for phylogenies labeled with the Cas9 indel LM distribution, it underperformed in distance-based tree reconstruction with only eight LMs due to increased homoplasy. Consequently, only the neighbor joining and UPGMA methods were used to reconstruct phylogenies generated with PEtracer LM data. Hybrid approaches, which pair a greedy algorithm for solving the top of the tree with a more accurate method, such as integer linear programming, for solving the bottom of the tree, can perform well but were not evaluated here since they require high-confidence greedy splits (*72*).

Evaluation of simulated tree reconstruction

Two complementary metrics were used to evaluate the similarity between simulated ground-truth trees and reconstructed trees with UPGMA, neighbor joining, and Greedy tree reconstruction algorithms. The first metric was RF distance, which quantifies topological differences by counting the number of node partitions unique to each tree (*73*). The RF metric was normalized by the total number of partitions, resulting in a range between 0 and 1. As a distance metric, lower values indicate greater similarity.

The second metric was depth-normalized mean triplets correct, which compares trees based on the relative ordering of leaf triplets rather than the set of partitions. To compute this metric, 1000 triplets were sampled from the ground-truth tree, distributed evenly across LCA depths to minimize bias introduced by most triplets having an LCA near the tree's root. The fraction of triplets with conserved ordering in the reconstructed tree was then calculated with values that range from 0 to 1. As a similarity metric, higher values indicate greater similarity.

All simulations were repeated 10 times, with reconstruction accuracies and parameters for each iteration reported tables S1 to S3. The mean value for each metric is presented with ribbons indicating standard deviation where applicable. In each iteration, the tree topology was kept constant while the simulated tracing data varied based on the specified parameters, except when assessing the effect of tree size on reconstruction accuracy. For values presented in heatmaps, the performance of the best algorithm for the given parameter combination is reported.

LM distributions

The uniformity of LM installation probabilities was quantified using normalized entropy (H_{norm}), calculated with the following formula:

$$H_{\text{norm}} = -\frac{\sum_{i=1}^{S} p(\text{LM}_i)\log p(\text{LM}_i)}{\log(S)}$$

Here, $p(\text{LM}_i)$ represents the probability of the i-th LM being installed, and S is the total number of lineage marks.

In simulations, the LM distributions with a given H_{norm} were generated by adjusting the λ parameter of an exponential distribution until the target H_{norm} was achieved. Thus, the probability of the i-th LM is given by:

$$p(\text{LM}_i) \propto \lambda e^{-\lambda i}$$

with probabilities normalized to sum to 1. The CRISPR-Cas9 indel distribution used in these simulations was empirically derived from data published by Yang *et al.* 2022 (*54*).

Edit rate simulations

A simplified simulation framework was used to model the relationship between edit rate and edit saturation over time. In this framework, it was assumed that all cells proliferate uniformly at a rate of one division per day, with no cell death. Instead of modeling all cells, which is computationally expensive because the number of cells grows exponentially, we model 100 cells over the duration of the simulation. Each parameter combination was simulated 10 times, and the mean values were reported.

To determine the edit rates required to achieve 50, 60, 70, 80, and 90% saturation, simulations were performed for edit rates ranging from 0 to 0.3 edits/day in increments of 0.001. For each experiment length, spanning 1 to 100 days, the edit rate closest to achieving the desired saturation value was identified.

To identify the minimum number of ESs required for 70, 80, and 90% of cell divisions to be marked by an edit as a function of tree size, we simulated experiments from 1 to 30 days in length—corresponding to tree sizes between 2^1 and 2^{30} cells. For each tree size, the number of ESs was varied from 1 to 120 (in increments of 1), and edit rates were tested from 0 to 0.3 (in increments of 0.001). The minimum number of ESs required to achieve the target edit fraction was determined for each tree size using the optimal edit rate for each experiment length. These values represent a lower bound on the required number of ESs, as the simulations assume uniform division rates, perfect detection of all cells in the tree, and ideal tracing kinetics. Results from these simulations are reported in table S4.

General mammalian cell culture conditions

4T1 cells (American Type Culture Collection (ATCC) CRL-2539) were maintained in RPMI 1640 medium (Thermo Fisher Scientific) supplemented with 10% FBS and 1% Penicillin-Streptomycin-Glutamine (PSG) (Thermo Fisher Scientific). B16-F10 cells (ATCC CRL-6475) were maintained in Dulbecco's Modified Eagle's Medium (DMEM; Thermo Fisher Scientific) supplemented with 10% FBS and 1% PSG. HEK293T (ATCC CRL-3216) and HEK293T/17 cells (ATCC CRL-11268) cells were maintained in DMEM supplemented with 10% FBS and 1% PSG. All cells were maintained at 37°C and 5% CO_2, authenticated by their suppliers, and tested negative for mycoplasma.

Mice

All mouse experiments described in this study were approved by the Massachusetts Institute of Technology Institutional Animal Care and Use Committee (IACUC, protocol numbers 0421-043-24 and 2311000598). To mitigate rejection of engineered 4T1 PEtracer cells, which are syngeneic in the Balbc/J background and express immunogenic components including PEmax (spCas9(H840A)–M-MLVRT) and mCherry (linked to lineage tracing cassettes), experiments were performed in F1 Balbc/J x SELECTIV (*143*) (LSL-Aavr-spCas9-mCherry) or Balbc/J x [SELECTIV x CMV-Cre (Aavr-spCas9-mCherry)] hybrids. Balbc/J (JAX #000651), SELECTIV (JAX #037553), and CMV-Cre (JAX #006054) mice were procured from JAX and bred in house. Tumor experiments were performed in female animals between 8 and 12 weeks of age. We performed tail vein lineage tracing experiments in three mice and

11 total tumors to capture biological variability. Statistical tests were not used to determine sample size and investigators were not blinded to experimental groups.

Cloning of tracing cassette and editor constructs

DNA fragments for cloning were either ordered from Integrated DNA Technologies (IDT) or Twist Bioscience as dsDNA fragments or PCRed using Phusion Green Hot Start II High-Fidelity PCR 2x Master Mix (Thermo Fisher Scientific) with the following cycle conditions: 98°C for 2 min, [98°C for 10 s, Primer specific Tm for 20 s, 72°C for 30 s/kb of amplified insert]x34, 72°C for 2 min. GC rich sequences like universal chromatin opening elements (UCOEs) were amplified using the KOD Xtreme Hot Start polymerase for 34 cycles following manufacturer's protocols (Sigma Aldrich, 71975-M). All PCR-amplified products were DpnI treated following manufacturer's protocols (New England Biolabs, NEB). DpnI-treated fragments were run out on either 1 or 2% agarose gels (VWR) supplemented with 1 ul of ethidium bromide (Thermo Fisher Scientific) per 10 ml of 1x Tris-Acetate-EDTA (TAE) buffer. Amplified DNA was extracted from bands excised from the gel using the QIAquick Gel Extraction Kit (Qiagen). Gibson assembly was used to assemble fragments into completed plasmids according to manufacturer's instructions. Specifically, we use 5 ul of NEBuilder 2x HiFi DNA Assembly master mix (NEB) combined with 0.05 pmol backbone and 0.15 pmol of each insert fragment brought to a final volume of 10 ul with ultrapure nuclease free water. Assemblies were incubated for 60 min at 50°C prior to transformation into NEB stable competent *E. coli* (C3040, NEB). Transformation included a 15-min incubation of 50 ul of competent cells and 5 ul of assembly mix prior to a 30-s heat shock at 42°C and a subsequent 5-min incubation on ice before plating on 100 μg/ml carbenicillin-containing agar plates. The Illustra TempliPhi 100 amplification kit (Cytivia) was used to perform rolling circle amplification (RCA) on bacterial colonies prior to Sanger sequencing or Whole Plasmid Sequencing (WPS) (Quintara Biosciences). Plasmid DNA was isolated from colonies with the correct sequence using the PureYield Plasmid Miniprep system (Promega) or with the ZymoPure II plasmid Midiprep system (Zymo Research).

The following tracing cassette, editor constructs, and other plasmids used in this manuscript were assembled using this protocol and are listed here for ease of reference. The sequence features of plasmids are described below. Addgene plasmid numbers are provided in parentheses where appropriate.

- Lentiviral PEmax editor that includes 5′-UCOE- EF1a short promoter (EFS)-PEmax-P2A-GFP-KASH-WPRE (Addgene #238541)
- Lentiviral lineage tracing cassettes that includes 5′-UCOE- EF1a promoter- mCherry- T7 promoter- T3 promoter- PacI digestion site-183-nt MERFISH decodable integration barcode (intBC)- Forward primer for sequencing assays-30-nt sequencing intBC- PacI digestion site- Edit Site 1 and flanking sequences- Edit Site 2 and flanking sequences- Edit Site 3 and flanking sequences- Reverse primer for sequencing assays- BGH polyA-500-nt common sequence (Addgene pooled library #238548)
- PiggyBac lineage tracing cassettes that includes 5′-UCOE- mCherry- T7 promoter- T3 promoter- PacI digestion site-183-nt MERFISH decodable intBC- Forward primer for sequencing assays- 30-nt sequencing intBC- PacI digestion site- Edit Site 1 and flanking sequences- Edit Site 2 and flanking sequences- Edit Site 3 and flanking sequences- Reverse primer for sequencing assays- SV40 polyA-500-nt common sequence
- Lentiviral puromycin static barcoding construct that includes 5′-UCOE- EF1a promoter-puromycin resistance cassette-10 N random barcode-BGH polyA (Addgene pooled library #238546)
- Lentiviral blasticidin static barcoding construct that includes 5′-UCOE- EF1a promoter-blasticidin resistance cassette-10 N random barcode-BGH polyA (Addgene pooled library #238547)
- Lentiviral CROP-seq BFP epegRNA expression vector that includes 5′-EF1a promoter- BFP- WPRE- Split 3′ LTR- hU6 promoter- pegRNA- Split 3′ LTR
- Three BsmBI-digestible PiggyBac 8-mer pegArray acceptor vectors and one PaqCI-digestible 24-mer pegArray acceptor vector that includes 5′-UCOE- EF1a promoter-BFP-WPRE-SV40 polyA-Digestion Cassette for pegArray insertions (ES1 8-mer vector = Addgene plasmid #238542, ES2 8-mer vector = Addgene plasmid #238543, ES3 8-mer vector = Addgene plasmid #238544, 24-mer vector = Addgene plasmid #238545).

Cloning of individual pegRNAs or epegRNAs

Individual pegRNAs and epegRNAs were cloned using Gibson assembly as described above. Briefly, hU6-pegRNA-GG-acceptor backbone (Addgene, #132777) expressing a guide RNA under the human U6 promoter was digested using BsaI-HFv2 (NEB, R3733L) according to the manufacturer's protocol. Digested backbone gel extracted as described above from a 1% agarose gel. eBlock fragments were ordered from IDT or Twist Bioscience encoding the relevant pegRNA or epegRNA sequences which were resuspended in ultrapure nuclease-free water. Gibson assemblies, transformations, colony picking, and DNA isolation from colonies were carried out as described above.

Testing of individual pegRNAs or epegRNAs

Individual pegRNAs or epegRNAs were evaluated with transfections using low passage HEK293T cells plated into tissue culture-treated (TC-treated) 96-well flat bottom dishes (Corning, CLS3799). Cells were maintained in TC-treated 15cm dishes. For transfections, media was removed, cells were rinsed once with sterile 1x phosphate-buffered saline (PBS) (Thermo Fisher Scientific), and trypsinized with 0.25% trypsin (Gibco). Following a 5-min incubation at 37°C for 5 min to ensure complete trypsinization, cells were resuspended in DMEM + 10% FBS to inactivate the trypsin and carefully triturated prior to cell counting using the Countess II automated cell counter (Thermo Fisher Scientific). Cells were diluted to 18,000 cells/ml and 100 ul of diluted cells were plated into 96-well dishes. Cells were transfected 18 to 24 hours after plating when they were ~60 to 70% confluent. For each well, transfections were carried out by combining (i) a DNA mixture comprising 200 ng of prime editor expressing plasmid (pCMV-PE2 plasmid = Addgene plasmid #132775; pCMV-PEmax plasmid = Addgene plasmid #174820) and 50 ng of a given pegRNA or epegRNA plasmid resuspended in 5 ul of OptiMEM (Thermo Fisher Scientific) with (ii) 0.5 ul of Lipofectamine 2000 (Thermo Fisher Scientific, 11668019) and 4.5 ul of OptiMEM (Thermo Fisher Scientific, 31985062). This mixture of DNA and lipid were mixed and incubated at room temperature for 15 min prior to adding dropwise to cells. At 72 hours posttransfection, media was removed, cells were washed with 1x PBS, and genomic DNA (gDNA) was extracted by adding 35 ul of freshly-prepared lysis buffer [10 mM Tris-HCl pH 7.5, 0.05% sodium dodecyl sulfate (SDS, Thermo Fisher Scientific), 25 μg/ml Proteinase K (Thermo Fisher Scientific)] to each well prior to a 1-hour incubation at 37°C. Following this digestion, gDNA was transferred to a 96-well PCR plate (Thermo Fisher Scientific) and incubated on a deep-well PCR machine for 30 min at 80°C to inactivate the Proteinase K; we term this isolation protocol the Fast Lysis procedure.

Sequencing and analysis of individual pegRNA and epegRNA editing outcomes

Genomic DNA isolated from HEK293T cells was amplified with two successive rounds of PCR as done previously (*66*). Briefly, PCR1 was conducted using a pool of forward and reverse primers where each primer was pooled in an equimolar fashion to enable sequencing on an Illumina MiSeq with minimal need for added PhiX. PCR1 was performed using 2 ul of Fast Lysis-isolated genomic DNA from HEK293T cells. Amplifications were carried out using Phusion Green Hot Start II High-Fidelity PCR 2x Master Mix with the following cycle conditions for the *EMX1*, HEK3, and *RNF2* endogenous genomic loci: 98°C for 2 min, [98°C for 10 s, 65°C for 20 s, 72°C for 1 min]x30, 72°C for 2 min. PCR1 amplicons were all checked on 2% agarose (VWR) gels with 1 ul of ethidium

bromide (Thermo Fisher Scientific) per 10 ml of 1x Tris-Acetate-EDTA (TAE) buffer. PCR2 amplification was carried out using TruSeq adapter PCR primers (listed in table S28) with the following conditions: 98°C for 2 min, [98°C for 10 s, 61°C for 20 s, 72°C for 1 min]x8, 72°C for 2 min. Two microliters of PCR1 were used as template DNA for each PCR2 reaction and 500 nM final primer concentrations were used. Following PCR2, all reactions for a given genomic locus were pooled together, run on a 2% agarose gel, and gel extracted using the QIAquick gel extraction kit (Qiagen). Concentrations of purified libraries were determined using a Qubit double-stranded DNA high sensitivity Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. Libraries were diluted to 4 nM and sequenced on a MiSeq (Illumina) using an Illumina MiSeq v2 Reagent kit.

Samples were demultiplexed with the MiSeq Reporter (Illumina). CRISPResso2 (v2.2.6) (*144*) was used to analyze demultiplexed reads. For all prime editing yield quantification, CRISPRessoBatch was run with default settings using the unedited allele as the reference amplicon and desired allele as Amplicon1 (-a parameter) and a 10-nt quantification window on each side of the cleavage position for a total length of 20 nt (-w parameter). Editing yield was calculated as: (number of Amplicon1-aligned reads) per (total reads aligned to all amplicons). For all experiments, indel yields were calculated as: (number of indel-containing reads) per (total reads aligned to all amplicons).

HEK Site *3* Genomic Locus Forward Primer Pool: ACACTCTTTCCCTACACGACGCTCTTCCGATCT-[1N to 4N]-ATGTGGGCTGCCTAGAAAGG

HEK Site *3* Genomic Locus Reverse Primer: TGGAGTTCAGACGTGTGCTCTTCCGATCTCCCAGCCAAACTTGTCAACC

***RNF2* Genomic Locus Forward Primer Pool:** ACACTCTTTCCCTACACGACGCTCTTCCGATCT-[1N to 4N]-ACGTCTCATATGCCCCTTGG

***RNF2* Genomic Locus Reverse Primer:** TGGAGTTCAGACGTGTGCTCTTCCGATCTACGTAGGAATTTTGGTGGGACA

***EMX1* Genomic Locus Forward Primer Pool:** ACACTCTTTCCCTACACGACGCTCTTCCGATCT-[1N to 4N]-CAGCTCAGCCTGAGTGTTGA

***EMX1* Genomic Locus Reverse Primer:** TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCGTGGGTTTGTGGTTGC

Cloning of comprehensive 1024 5-nt LM epegRNA libraries

Gibson assemblies between PCR amplified epegRNA backbones and short oligonucleotide libraries ordered from IDT were used to generate libraries representing all possible 5-nt LM libraries for each of the three target sequences. First, epegRNA backbones with optimized primer binding sequence and reverse transcription template (RTT) lengths were amplified as described in "Cloning of tracing cassette and editor constructs" using Phusion Green Hot Start II High-Fidelity PCR Master Mix with the primers listed below and 34 cycles of amplification. Two independent 5N LM libraries were ordered from IDT for each site, a top and bottom strand library. Libraries were diluted to 1 μM and 1 ul was used for each of 10 parallel Gibson assemblies with 0.05 pmol of amplified backbone for their respective target loci. All 100 ul of Gibson assembly product were cleaned up using PCR cleanup columns (Qiagen) to remove salts prior to electroporation into MegaX cells as previously described (*145*). Colonies were scraped into a combined pool, grown out for 8 hours in 100 ml terrific broth (TB) containing 100 μg/ml carbenicillin, and midi-prepped using the ZymoPure II Plasmid Midiprep Kit (Zymo Research) following manufacturer's procedures.

Forward primer for all: CGCGGTTCTATCTAGTTACGCGTTAAAC

***EMX1* reverse primer:** CTCCCATCACATGCACCGAC

HEK*3* reverse primer: TGATGGCAGAGGAAAGGAAGCC

***RNF2* reverse primer:** CTGAGGTGTTCGTTGCACCG

***EMX1* *5*N library top:** GTCGGTGCATGTGATGGGAGNNNNNTTCTTCTGCTCGGACGCGGTTCTATCTAGTTACG

***EMX1* *5*N library bottom:** CGTAACTAGATAGAACCGCGTCCGAGCAGAAGAANNNNNCTCCCATCACATGCACCGAC

HEK*3* *5*N library top: CTTCCTTTCCTCTGCCATCANNNNNCGTGCTCAGTCTGCGCGGTTCTATCTAGTTACG

HEK*3* *5*N library bottom: CGTAACTAGATAGAACCGCGCAGACTGAGCACGNNNNNTGATGGCAGAGGAAAGGAAG

***RNF2* *5*N library top:** CGGTGCAACGAACACCTCAGNNNNNGTAATGACTAAGATGCGCGGTTCTATCTAGTTACG

***RNF2* *5*N library bottom:** CGTAACTAGATAGAACCGCGCATCTTAGTCATTACNNNNNCTGAGGTGTTCGTTGCACCG

Testing of 1024 LM epegRNA libraries

Testing of 1024 LM epegRNA libraries for three test sites was carried out with transfections using low passage HEK293T cells plated into TC-treated six-well flat bottom dishes (Corning, 353046). Cell maintenance and splitting via trypsinization were conducted identical to 96-well format transfections. Following cell counting, cells were resuspended to 400,000 cells/ml and 1 ml of cells was added to each well of a six-well TC-treated plate to which an additional 1 ml of media was added to avoid evaporation. 18 to 24 hours after plating, three wells were transfected for each 1024 LM library to ensure comprehensive coverage of the entire library and biological consistency. Each library was transfected as before by combining a mixture of DNA with a lipid mixture prior to a 10-min incubation at room temperature prior to adding dropwise to six-well plates. For six-well formats DNA mixtures comprise 6 μg of PE2 plasmid and 1 μg of 1024 LM library was resuspended in a final volume of 25 ul of OptiMEM while lipid mixtures comprise 15 ul of Lipofectamine 2000 and 10 ul of OptiMEM. At 72 hours posttransfection, media was removed, cells were washed with 1x PBS, and trypsinized as above to isolate cell pellets. Genomic DNA was extracted using the DNeasy blood and tissue kit (Qiagen) following manufacturer's protocols.

Sequencing and analysis of 1024 LM epegRNA libraries

Plasmid DNA sequences were verified to ensure coverage of all edits 1024 LMs across the library for each of the three loci. Two successive rounds of PCR were used to generate plasmid libraries using primers listed below. For plasmid libraries, PCR1 was performed using Phusion Green Hot Start II High-Fidelity PCR 2x Master Mix with the following cycle conditions: 98°C for 2 min, [98°C for 10 s, 65°C for 20 s, 72°C for 1 min] x the number of cycles was empirically determined for each sample by qPCR, 72°C for 2 min. Ten reactions for each plasmid library where 500 ng of the plasmid library was used for each 25-ul amplification reaction along with 500 nM primer mix.

epegRNA Plasmid Library PCR1 Forward primer pool: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-[1N to 4N]-GAAAAAGTGGCACCGAGTCGG

epegRNA Plasmid Library PCR1 Reverse primer pool: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-[1N to 4N]- TTTGTGATGCTCGTCAGGGG

For genomic DNA, 48 25-ul PCR1 reactions with 500 ng of gDNA and 500 nM primer mix were run to ensure broad coverage across all samples. Genomic loci primers are the same as used previously for these amplifications ("Sequencing and analysis of individual pegRNA and epegRNA editing outcomes"). Amplifications were performed using Phusion Green Hot Start II High-Fidelity PCR 2x Master Mix with the following cycle conditions: 98°C for 2 min, [98°C for 10 s, 65°C for 20 s, 72°C for 1 min]x30, 72°C for 2 min. As before, longer extension times were used to ensure complete extension for each cycle. PCR1 amplicons were all checked on 2% agarose as above.

PCR2 amplification for both plasmid libraries and genomic DNA from transfected cells was carried out using Nextera adapter PCR primers (listed in table S29) with the following conditions: 98°C for 2 min, [98°C for 10 s, 61°C for 20s, 72°C for 1 min]x8, 72°C for 2 min. 5 ul of PCR1 products were used for each PCR2 reaction and each of the reactions was independently barcoded in PCR2. Following PCR2, all reactions for each genomic DNA or plasmid sample were pooled together, run on a 2% agarose gel, and gel extracted prior to quantification, library generation, and sequencing as described above

("Sequencing and analysis of individual pegRNA and epegRNA editing outcomes").

CRISPResso2 (v2.2.6) was used to analyze demultiplexed reads. For both input plasmid libraries and genomic DNA from edited cells, CRISPRessoBatch was run with default settings using the unedited allele as the reference amplicon and desired allele with a 5 N insertion as Amplicon1 (-a parameter) and a 10-nt quantification window on each side of the cleavage position for a total length of 20 nt (-w parameter). Frequencies of aligned reads were extracted from the output Alleles_frequency_table files. Aligned reads were trimmed to 25-nt flanking the cleavage position (−10 nt and +10 nt from the cleavage position to capture 5-nt insertions +10 nt of additional flanking sequence for edited alleles) and counts for reads with expected 5-nt insertions were quantified. On average, 52% (range 43 to 61%) of aligned reads from each gDNA sample matched an expected 5-nt insertion and 43% (range 30 to 55%) of reads matched the unedited sequence, with only 4.6% (range 1.7 to 8.2%) of reads representing unintended alleles (i.e., indels or sequencing/PCR errors). For the input plasmid libraries, on average 88% (range 79 to 93%) of reads matched the expected epegRNA sequence with a 5-nt insertion. For subsequent analysis, only aligned reads matched an expected 5-nt insertion were considered. Normalized counts per million values were $\log_2$ transformed, replicates averaged, and used to calculate the fold change for each 5-nt insertion sequence in the edited gDNA relative to the input plasmid library. The 5-nt insertion sequences were ranked by $\log_2$ fold change (see table S7) to nominate the top-performing LMs at each test site.

Cloning and testing of top-performing LMs at three test sites

The top 94, 96, and 85 LMs for the HEK3, *EMX1*, and *RNF2* loci were tested individually. epegRNAs for each of these LMs were cloned, plasmid DNA was isolated, transfected, sequenced and analyzed as described above ("Testing of individual pegRNAs or epegRNAs" and "Sequencing and analysis of individual pegRNA and epegRNA editing outcomes").

Generation and testing of orthogonal ES sequences

To make ES sequences orthogonal to the human genome as well as the genomes of other widely used model organisms, we reasoned that by making base changes in the 3-nt seed region of our ES sequences that we could preserve the primer binding site sequence for each ES while largely maintaining the base identity and annealing properties of the RTT to generate novel ES sequences that are not directly present in various genomes of interest. We generated two orthogonal sequence variants per ES. We designed protospacer sequences where we used the opposite Watson-Crick-Franklin pairing DNA base at each position from the endogenous human locus for the 3-nt genomic seed for Ortho v1 (i.e., if the ES sequence were 5′-GCT-3′ we would use 5′-CGA-3′), and used the reverse complement of the 3-nt seed for Ortho v2 (i.e., if the sequence were 5′-GCT-3′ we would use 5′-AGC-3′).

Sequencing and analysis of orthogonal ES sequences

To experimentally evaluate our ES orthogonalization strategies, we generated synthetic cassettes that harbored each of the *RNF2*, HEK3, and *EMX1* ES sequences and necessary flanking sequences to support efficient prime editing for the original, unmodified genomic ES sequences and for each of the two new orthogonalized ESs. These ESs were ordered as e-blocks from IDT and cloned using Gibson assembly into LTC vectors. This vector introduces ES sequences into the 3′ UTR of an mCherry reporter gene. These alleles were integrated into HEK293T cell genomes and cells harboring these constructs were isolated by sorting for mCherry+ cells using FACS 2 weeks after integration into the genome (representative sorting gates shown in fig. S16). Cells were subsequently transfected with epegRNAs targeting the corresponding sequences, cloned as described above ("Cloning of individual pegRNAs or epegRNAs").

Genomic DNA was isolated using the Fast Lysis procedure described for sequencing HEK293T genomic loci 72 hours posttransfection. Two successive rounds of PCR were used to amplify both synthetic ESs (PCR1 primers listed below) as well as endogenous genomic loci as described above (PCR1 primers listed in "Sequencing and analysis of individual pegRNA and epegRNA editing outcomes") with 2 ul of gDNA and 500 nM primer mix using the Phusion Green Hot Start II High-Fidelity PCR 2x Master Mix with the following cycle conditions: 98°C for 2 min, [98°C for 10 s, 65°C for 20 s, 72°C for 30 s]x28, 72°C for 2 min. PCR1 amplicons were all checked on 2% agarose as above. PCR2 amplification was carried out using Nextera adapter PCR primers with the following conditions: 98°C for 2 min, [98°C for 10 s, 61°C for 20s, 72°C for 30 s]x8, 72°C for 2 min. 5 ul of PCR1 products were used for each PCR2 reaction and each of the reactions was independently barcoded in PCR2. All steps through sequencing using the Illumina Miseq v2 Reagent kit were performed as above ("Sequencing and analysis of individual pegRNA and epegRNA editing outcomes"). Experiments to re-test optimal LMs for orthogonalized ES sequences were carried out using this same protocol.

LTC Forward primer pool: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-[1N to 4N]- GAATCCAGCTAGCTGTGCAGC

LTC Reverse primer pool: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-[1N to 4N]- CCTTAGCCGCTAATAGGTGAGC

CRISPResso2 (v2.2.6) was used to analyze demultiplexed reads. For all PE yield quantification, CRISPRessoBatch was run with default settings using the unedited allele as the reference amplicon and desired allele as Amplicon1 (-a parameter) and a 10-nt quantification window on each side of the cleavage position for a total length of 20 nt (-w parameter). Editing yield was calculated as: (number of Amplicon1-aligned reads) per (total reads aligned to all amplicons). For all experiments, indel yields were calculated as: (number of indel-containing reads) per (total reads aligned to all amplicons).

In silico characterization of orthogonalized ES sequences

To further ensure compatibility of Ortho v1 ES sequences for use in widely used model systems, we used the Cas-OFFinder in silico prediction algorithm (*86*) (v2.4.1) to identify potential CRISPR/Cas9-dependent off-target loci with three or fewer mismatches or bulges to our three OrthoV1 protospacers of interest. Cas-OFFfinder nominated 69, 174, 13, and 4 Cas9-dependent off-targets with 3 or fewer mismatches or bulges (MMB) for *H. sapiens*, *M. musculus*, *D. rerio*, and *D. melanogaster* genomes, respectively. There were no predicted genomic off-targets with 1 mismatch in any of the analyzed genomes (fig. S2D). All predicted 2-mismatch off-targets were within intergenic or intronic regions of the genome with the exception of one in the promoter of the *apoea* gene in the *D. rerio* genome. Among the 251 putative genomic off-targets with 3 mismatches to our sgRNAs, only 4 were within exons of genes—2 each in the human and mouse genomes—and 14 were in other regulatory genomic elements. When each of these 18 putative off-target loci were examined, we did not identify homologous sequences within 100nt downstream of the protospacer likely to support prime editing.

LM detection cross-hybridization calculations

To select lineage marks suitable for accurate hybridization probe readout, we used NUPACK (v4.0.1.9) (*146*) simulations to evaluate off-target binding for various LM sequence combinations. Each simulation involved hybridizing an equimolar mixture of DNA readout sequences (20-nt complementary region) with an equimolar mixture of DNA ES sequences containing the selected LMs. Simulations were conducted using a 100-fold excess of readout sequences (10 nM versus 0.1 nM) at 43°C in a sodium ion concentration of 0.3 M. For each LM sequence, the free-energy difference (ΔG) between correct and incorrect probe binding was estimated. Cross-hybridization was quantified as the fraction of ES molecules bound by incorrect readout sequences. Separate simulations were performed for the Ortho v1 Edit Site 1, Edit Site 2,

and Edit Site 3 sequences. Results from these simulations are reported in table S8.

To investigate the effect of LM insert length on probe cross-hybridization, we simulated mixtures of ES sequences containing random LM inserts of varying lengths (2 to 8 nt). For each LM length, simulations were repeated 10 times with each iteration using eight random LM sequences of the specified length plus the unedited sequence while keeping the total complementarity length and other hybridization parameters constant. The mean fraction of correct probes binding each ES (1 – cross-hybridization fraction) is shown in fig. S1J and results for each iteration are reported in table S4.

Design and cloning of LTC libraries

With three ES sequences selected, we finalized our LTC design. These cassettes consist of the following sequence organization: 5′-UCOE-EF1a promoter- mCherry- T7 promoter- T3 promoter- PacI digestion site- 30-nt sequencing barcodes- sequencing primer- MERFISH decodable intBC- Forward primer for sequencing assays- sequencing intBC- PacI digestion site- Edit Site 1 and flanking sequences- Edit Site 2 and flanking sequences- Edit Site 3 and flanking sequences- Reverse primer for sequencing assays- polyA. For lentiviral construct designs, we included 500 nt of constant sequence downstream for optimizing imaging readout experiments and used a BGH polyA sequence. For PiggyBac transposase-compatible vectors, we introduced this 500-nt constant sequence before an SV40 polyA.

For testing of 24-mer pegArrays and epegRNA mismatch libraries, we generated a PiggyBac lineage tracing cassette plasmid library with a random 10-nt sequencing barcode (LTCv0).

For in vitro single cell lineage tracing experiments, we generated a lentiviral LTC plasmid library with 125-nt MERFISH barcodes with an additional 30-nt sequencing barcode (LTCv1). Specifically, 125-nt MERFISH barcodes were generated by choosing 6 among 21 of previously published MERFISH readout sequences (*147*) (B37 to B66, 20 nt for each) and concatenating them together with an "A" between each pair of readout sequences. Thirty-nucleotide sequencing barcodes were randomly generated.

For the in vitro imaging experiments and all in vivo experiments, we generated a lentiviral LTC plasmid library with matching 30-nt sequencing barcodes and 183-nt MERFISH barcodes (LTCv2). To design these orthogonal barcodes, we first randomly selected 9 of the 25-nt orthogonal DNA among 240,000 published DNA barcodes (*148*), concatenated them in a randomly permuted order and selected the best permutation by checking if all 15-mer are exclusive to this barcode. We repeated this process to select up to 6000 candidate barcodes. We then predicted the self-binding partition function by "pfunc" from Nupack (*146*) and selected the candidate barcodes with highest free energy to minimize self-binding. After filtering, we inserted the sequencing primer at position 30 of candidate barcodes and designated the first 30 nt as sequencing barcodes, and then selected the next 183 nt as MERFISH barcodes. We concatenated the 5′ sequence (T7, T3 and PacI site) and 56 nt of 3′ sequence (Edit Site 1) together as candidate intBCs for cross-hybridization testing. We first partitioned each 183-nt MERFISH barcode into six non-overlapping 30-nt segments, and used the reverse complement of these segments as candidate probes. To evaluate probe cross-hybridization, we enumerated the pairs of candidate intBCs, generated a Nupack Tube class with the candidate probes and these two candidate intBCs, and predicted the free-energy and partition function for all possible bindings. After prediction, we summarized probe cross-hybridization for all candidate intBCs and retained intBCs with less than 1% of predicted average probe off-targeting probability. For barcodes that passed these requirements, we further filtered to have no 17 nt matching to human, mouse, zebrafish and *Drosophila* genomes.

As such, we ordered libraries of 2171 intBCs (table S6) from Twist Bioscience and amplified this pooled library using Phusion Green Hot Start II High-Fidelity PCR 2x Master Mix with the following cycle conditions: 98°C for 2 min, [98°C for 10 s, 65°C for 20 s, 72°C for 1 min] where the number of cycles for amplification were determined by quantitative PCR, followed by 72°C for 2 min. Given that libraries can be provided with variable quality, this qPCR determination was important for ensuring library quality. Following library amplification, samples were cleaned up with a QIAquick PCR purification column (Qiagen) according to manufacturer's protocols. These amplified DNA fragments were assembled into PacI-digested and Quick CIP-treated (NEB) acceptor backbone via Gibson assembly using NEBuilder 2x HiFi DNA Assembly master mix as large scale 100-ul reactions described above. The entire 100 ul of Gibson assembly product was cleaned up using PCR cleanup columns (Qiagen) to remove salts prior to electroporation into MegaX cells as previously described. Colonies were scraped into a combined pool, grown out for 8 hours in 500-ml terrific broth (TB) containing 100 μg/ml carbenicillin, and midi-prepped using the ZymoPure II Plasmid Midiprep Kit (Zymo Research) following manufacturer's procedures.

LTCv2 Cloning Forward primer: AATTAACCCTCACTAAAGGGATAATTTAATTAA

LTCv2 Cloning Reverse primer: GTCGTAATGACTAAGATGACTGCCATTAATTAA

Sequence validation and analysis of LTC libraries

Lentiviral plasmid libraries were sequence confirmed as described for the 1024 epegRNA libraries with modest differences. As before, plasmid libraries were amplified with two successive rounds of PCR as done to sequence genomic loci. PCR1 was performed using Phusion Green Hot Start II High-Fidelity PCR 2x Master Mix with the following cycle conditions: 98°C for 2 min, [98°C for 10 s, 65°C for 20 s, 72°C for 1 min] where the cycle number was determined by qPCR, followed by 72°C for 2 min, and a hold at 12°C. PCR2, gel extraction, library preparation, and sequencing were carried out as before.

intBC Sequencing Validation Forward primer: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-[1N-4N]-CATCGATACCTAATACGACTCACTATAGGGAGAG

intBC Sequencing Validation Reverse primer: GTCTCGTGGCTCGGAGATGTGTATAAGAGACAGNCCTGTCGTAATGACTAAGATGACTGCC

To assess the balance of library elements, we used the FastX toolkit (v0.0.14) to mask low quality bases (<q20) and collapse reads. We removed common flanking sequences and summed reads matching the whitelist of intBC barcodes for sequences with Hamming distance <4 to account for sequencing errors and the length of intBC sequences.

Plasmid design and cloning strategy for 8-mer and 24-mer pegArrays

To clone 24-mer arrays comprising an 8-mer or each of the ESs, we designed a two-step Golden Gate assembly cloning strategy. For the first step, 8-mer arrays are cloned into three distinct acceptor backbones, one for each ES where the first cloning step is carried out with the Type IIS restriction enzyme BsmBI (NEB, R0739L). These backbones can then be used for a second cloning step using the Type IIS restriction enzyme PaqCI (NEB, R0745L) for the final 24-mer assembly. The overhangs for the 8-mer assembly are the same for each of 8-mer backbones. The backbone overhangs are 5′-GGAG-3′ and 5′-CTAA-3′. The intervening 8 epegRNA inserts use the following overhangs: 5′-GGAG-INS 1-TGCC-3′, 5′-TGCC-INS 2-GCAA-3′, 5′-GCAA-INS 3-ACTA-3′, 5′-ACTA-INS 4-TTAC-3′, 5′-TTAC-INS 5-CAGA-3′, 5′-CAGA-INS 6-TGTG-3′, 5′-TGTG-INS 7-GAGC-3′, 5′-GAGC-INS 8-CTAA-3′. Fragments with these overhangs were ordered from either IDT or Twist Bioscience using table S30 as a design guide. The final 24-mer array has the three 8-mers in the following order to optimize for LM balance: Edit Site 2, Edit Site 3, then Edit Site 1. To generate 24-mers, sequence verified 8-mer arrays were digested with PaqCI to generate 8-mer array fragments with the following overhangs: 5′-GGAG-[Edit Site 2 8-mers]-TGTG-3′, 5′-TGTG-[Edit Site 3 8-mers]-GAGC-3′, 5′-GAGC-[Edit Site 1 8-mers]-CTAA-3′. These

fragments could then be cloned into a 24-mer acceptor backbone digested with PaqCI. All golden gate acceptor plasmids are provided on Addgene.

Cloning protocol for 8-mer pegArrays

Three different acceptor backbones were generated for 8-mer assemblies, one for each of the three ESs. These acceptor backbones include an EF1a promoter driving a BFP reporter gene and have BsmBI type-IIS restriction cut sites. DNA fragments were ordered from Twist Bioscience that include a U6 Pol III promoter upstream of epegRNAs for each of the eight LMs for a given ES. Golden Gate assemblies were performed using 3 nM of each U6–epegRNA fragment (typically 1 ul of each ~500-nt fragment diluted to 20 ng/ul), 1.1 nM of the relevant backbone (typically 1 ul of the ~8200-nt BsmBI pre-digested backbone diluted to 25 ng/ul), 2 ul of NEBridge Golden Gate Assembly Kit (BsmBI-v2) (NEB, E1602L), 2 ul T4 ligase buffer, and 7 ul of nuclease free water. Assemblies were carried out following manufacturer's recommendations: 99 cycles of 5 min at 42°C and 5 min at 18°C before a 10-min 60°C incubation and subsequent 4°C hold. These reactions were transformed into NEB Stable Competent *E. coli* (NEB, C3040), incubating on ice for 20 min prior to a 30-s heat shock at 42°C. After heat shock, bacteria were incubated on ice for 5 min prior to plating on 100 mg/ml carbenicillin resistant agar plates. Plates were incubated overnight at 30°C prior to picking colonies into 5 ml of 100 mg/ml carbenicillin-containing terrific broth, outgrowth overnight at 30°C, and DNA isolation using the Promega PureYield DNA Miniprep System. Some bacteria were expanded into 100-ml cultures overnight while isolated plasmids were diluted to 100 ng/ul and submitted for whole plasmid sequencing. DNA was isolated from these larger cultures using the ZymoPure II Plasmid Midiprep Kit. Plasmids were sequence confirmed by whole plasmid sequencing and further validated to have the correct insert by gel electrophoresis to be the correct size following digestion with PaqCI.

Cloning protocol for 24-mer pegArrays

24-mer assemblies require four pieces of DNA that include a PaqCI digested 24-mer backbone and a PaqCI digested fragment for each ES 8-mer. PaqCI digestions were performed overnight using 4 ul of CutSmart buffer, 4 ul of PaqCI activator, 2 ul of PaqCI, and 30 ul of the plasmid of interest. These digested fragments were gel extracted using the QIAquick Gel Extraction Kit (Qiagen) following manufacturer's recommendations. Digested fragments were assembled into 30 ul PaqCI Golden Gate assemblies with 3.33 nM of the PaqCI digested 8-mer fragments (typically 1 ul of the ~4000-nt fragment diluted to 150 ng/ul), 1 ul of PaqCI pre-digested 24-mer acceptor backbone diluted to 25 ng/ul, 2.5 ul PaqCI enzyme, 1.25 ul PaqCI activator, 10 ul NEBridge Ligase Master Mix (M1100L), and water to bring the reaction to 30 ul total volume. Assemblies were carried out following manufacturer's recommendations, carrying out 99 cycles of 5 min at 37°C and 5 min at 16°C before a 10-min 60°C incubation and subsequent 4°C hold. All subsequent steps of transformation, plating, colony picking, and DNA prep and sequence confirmation were carried out as described for 8-mer assemblies above.

Testing of 8-mer and 24-mer pegArrays

We introduced PEmax editor and LTCs (LTCv0) into B16-F10 cells using sequential PiggyBac transposition. One million cells were nucleofected with 1.5 μg of transfer plasmid and 150 ng of Super PiggyBac Transposase Expression Vector (System Biosciences PB210PA-1) using the SF Cell Line 4D-Nucleofector Kit L (Lonza V4XC-3024) with program CM-130 and sorted for GFP+ (PEmax+) cells and GFP+/mCherry+ (PEmax+/lineage tracing cassettes+) at least 8 days after nucleofection to ensure stable expression. We then introduced 8-mer and 24-mer pegArray variants into GFP+/mCherry+ cells using the same nucleofection conditions and sorted GFP+/mCherry+/BFP+ (PEmax+/lineage tracing cassettes+/pegArray+) at least 8 days after nucleofection. Representative sorting gates shown in fig. S16. Genomic DNA was isolated using QuickExtract DNA Extraction Solution (Biosearch Technologies, QE09050) following the manufacturer's protocol.

Sequencing and analysis of 8-mer and 24-mer pegArrays

To assess pegRNA competition, LTCs were amplified and sequenced to quantify lineage mark installation efficiency. Two successive rounds of PCR were used to amplify lineage tracing cassettes with 2 ul of gDNA and 500 nM primer mix using the Phusion Green Hot Start II High-Fidelity PCR 2x Master Mix with the following cycle conditions: 98°C for 2 min, [98°C for 10 s, 65°C for 20 s, 72°C for 30 s]x28, 72°C for 2 min. PCR1 amplicons were all checked on 2% agarose as above. PCR2 amplification was carried out using Nextera adapter PCR primers with the following conditions: 98°C for 2 min, [98°C for 10 s, 61°C for 20 s, 72°C for 30 s]x8, 72°C for 2 min. Five microliters of PCR1 products were used for each PCR2 reaction and each of the reactions was independently barcoded in PCR2. Following PCR2, all reactions for each genomic DNA or plasmid sample were pooled together, run on a 2% agarose gel, and gel extracted prior to quantification, library generation, and sequencing as described above.

LTC Forward primer pool: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-[1N to 4N]- GAATCCAGCTAGCTGTGCAGC

LTC Reverse primer pool: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-[1N to 4N]- CCTTAGCCGCTAATAGGTGAGC

CRISPResso2 (v2.2.6) was used to analyze demultiplexed reads. CRISPResso was run with default settings using the unedited allele as the reference amplicon and desired allele with a 5N insertion as Amplicon1 (-a parameter) and a 10-nt quantification window on each side of the cleavage position for a total length of 20 nt (-w parameter). Frequencies of aligned reads were extracted from the output Alleles_frequency_table files. Aligned reads were trimmed to 25 nt flanking the cleavage position (−10 nt and +10 nt from the cleavage position to capture 5-nt insertions +10 nt of additional flanking sequence for edited alleles) and counts for reads with expected 5-nt insertions were quantified. For subsequent analysis, only aligned reads matched an expected 5-nt insertion were considered to calculate the relative editing efficiency for each pegRNA. Read counts for each 5nt insertion and array are reported in table S9.

Cloning of a comprehensive epegRNA mismatch library

We designed a library of epegRNA protospacer mismatches that includes all possible single letter variants at positions 2 to 20 of an SpCas9 protospacer (where the NGG PAM is 21 to 23). We preserved a guanine (G) at position 1 to ensure efficient transcription by the human U6 promoter. Inspired by the CROP-seq vector (*88*), we cloned this epegRNA library into the 3′ UTR of a lentiviral genome (modified from Addgene plasmid #86708 to replace puromycin with BFP and introduce BsmBI epegRNA cloning site). This vector ensures that the sequence of each epegRNA could be captured in droplet-based scRNA-seq experiments while producing an active copy of the epegRNA through the genome integration process. epegRNA protospacer mismatches were either synthesized and cloned by Twist Bioscience or ordered as eBlock fragments from IDT and cloned into the BsmBI-digested CROP-seq vector backbone using Gibson assembly. All protospacer mismatches are listed in table S10.

Lentiviral production

Lentiviral particles were generated as previously described (*149*). Briefly, HEK293T/17 cells (ATCC CRL-11268) were maintained in DMEM (Thermo Fisher Scientific) supplemented with 10% (v/v) FBS and 1% (v/v) PSG at 37°C with 5% CO_2. Cells were split 1:8 2 days prior to use to ensure rapid cycling. Twenty-three million cells were plated per 15-cm dish on the day prior to transfection. The next day, cells were ~90% confluent and were transfected with the following transfection

mix: 200 ul FuGENE HD transfection reagent and a DNA mixture with 1.3 pmol of psPAX2 (Addgene plasmid #12260), 0.72 pmol pMD2.G (Addgene plasmid #12259), and 1.64 pmol of the transfer plasmid. DNA and Fugene mix were resuspended in a final volume of 3 ml of OptiMEM serum free media. This mixture was vortexed gently and incubated for 15 min at room temperature before adding dropwise to cells. Six hours posttransfection, transfection media was removed and replaced with Opti-MEM I Reduced Serum Medium (OPTI-MEM) with GlutaMAX Supplement (Invitrogen, 31985088) supplemented with 5%FCS, 1 mM sodium pyruvate (Fisher Scientific), and 1x MEM nonessential amino acids (FisherScientific) (cOPTI-MEM) supplemented with 1x ViralBoost reagent (AlStem). Forty-eight hours posttransfection, media was collected, spun at 3000 g for 15 min to remove cell debris and concentrated using Lenti-X-Concentrator (TakaraBio, 631232) following the manufacturer's instructions. Concentrated virus was resuspended in OPTI-MEM in 1% of the original culture volume without supplements. Lentiviral particles were subsequently aliquoted and frozen at −80°C.

epegRNA mismatch library profiling to tune lineage tracing kinetics

We introduced LTCs where each of the three ESs were concatenated together (LTCv0) using PiggyBac transposition into two mismatch repair–competent cell lines, B16-F10 (a melanoma-derived line; ATCC CRL-6475) and 4T1 (a breast cancer-derived line; ATCC CRL-2539). We also introduced the PEmax editor using PiggyBac transposition and sorted for cells that were both GFP+ and mCherry+ to select for editor and tracing cassette containing cells, respectively. For PiggyBac transposition, 1 million cells were nucleofected with 1.5 μg of transfer plasmid and 150 ng of Super PiggyBac Transposase Expression Vector (System Biosciences, PB210PA-1). We used the SF Cell Line 4D-Nucleofector Kit L (Lonza V4XC-3024) with program CM-130 for B16-F10 cells and the SE Cell Line 4D-Nucleofector Kit L (Lonza V4XC-1024) with program CM-150 for 4T1 cells. We then treated GFP+/mCherry+ cells with lentivirus expressing the CROP-seq BFP epegRNA library, sorted BFP+ cells, and passaged these cells for 28 days, collecting samples throughout the duration of the experiment. Representative sorting gates shown in fig. S16.

scRNA-seq was performed on harvested cells to link epegRNA mismatches with lineage tracing editing rates. scRNA-seq libraries were prepared using the 10X Genomics Chromium Next GEM Single Cell 3′ Kit v3.1 following the manufacturer's instructions followed by targeted enrichment of CROP-seq epegRNA and LTC transcripts based on published protocols (*39*, *53*, *150*). To increase diversity in Read2 during sequencing, we used a pool of primers with an 8-nt stagger following the Illumina Read2 primer binding site.

Two successive rounds of PCR were used to prepare LTC libraries from amplified full-length cDNA. For the first PCR reaction (PCR1), 2 ul of amplified cDNA and 600 nM primer mix listed below (LTC_10X_PCR1_0-8_F, 10X_PCR1_R) were used. Amplifications were performed using KAPA HiFi HotStart ReadyMix with the following cycle conditions: 95°C for 3 min, [98°C for 20s, 65°C for 15s, 72°C for 15s]x19-25, 72°C for 1 min. PCR2 amplification was carried out using 4 ul of a 1:25 dilution of PCR1 product and 300 nM of the reverse primer (10X_PCR2_R) listed below and Nextera i7 adapter PCR primers (listed in table S29) with the following conditions: 95°C for 3 min, [98°C for 20s, 72°C for 30 s]x12, 72°C for 1 min.

scRNA-seq LTC amplification primers:

PCR1 forward (LTC_10X_PCR1_0-8_F): GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG -[1N to 8N]-GAATCCAGCTAGCTGTGCAGC

PCR1 reverse (10X_PCR1_R): ACACTCTTTCCCTACACGACG

PCR2 forward (Nextera i7): CAAGCAGAAGACGGCATACGAGATN-NNNNNNNGTCTCGTGGGCTCGG

PCR2 reverse (10X_PCR2_R): AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC

Three successive rounds of PCR were used to prepare epegRNA libraries from amplified full-length cDNA. For the first PCR reaction (PCR1), 2 ul of amplified cDNA and 300 nM primer mix listed below (pegRNA_10X_PCR1_F, pegRNA_10X_PCR1_R) were used. Amplifications were performed using KAPA HiFi HotStart ReadyMix with the following cycle conditions: 95°C for 3 min, [98°C for 20s, 65°C for 15s, 72°C for 15s]x29-33, 72°C for 1 min. PCR2 amplification was carried out using 1 ul of a 1:25 dilution of PCR1 product and 300 nM of the primer mix listed below (pegRNA_10X_PCR2_0-8_F, 10X_PCR2_R) listed below with the following conditions: 95°C for 3 min, [98°C for 20s, 65°C for 15s, 72°C for 15s]x19, 72°C for 1 min. PCR3 amplification was carried out using 4 ul of a 1:25 dilution of PCR2 product and 300 nM of the reverse primer (10X_PCR2_R) listed below and Nextera i7 adapter PCR primers (listed in table S29) with the following conditions: 95°C for 3 min, [98°C for 20s, 72°C for 30 s]x12, 72°C for 1 min.

scRNA-seq epegRNA amplification primers:

PCR1 forward (pegRNA_10X_PCR1_F): ACACTCTTTCCCTACACGACG

PCR1 reverse (pegRNA_10X_PCR1_R): TTTCCCATGATTCCTTC-ATATTTGC

PCR2 forward (pegRNA_10X_PCR2_0-8_F): GTCTCGTGGGCTCG-GAGATGTGTATAAGAGACAG-[1N-to-8N]-CTTGTGGAAAGGACG-AAACAC

PCR2/PCR3 reverse (10X_PCR2_R): AATGATACGGCGACCACCG-AGATCTACACTCTTTCCCTACACGACGCTC

Following PCR amplification, lineage tracing cassette and epegRNA libraries were purified and size-selected using SPRI magnetic beads (0.9x single-sided selection) and quantified by BioAnalyzer (Agilent) to assess the size and purity of final libraries and Qubit double-stranded DNA high sensitivity Kit (Thermo Fisher Scientific) to determine the concentrations. Libraries were diluted to 4 nM and sequenced on a Nextseq 2000 (Illumina). All PCRs were first monitored to determine optimal cycle number by qPCR by adding 0.6x SYBR Green (Thermo) to the PCR reactions.

scRNA-seq data processing

All scRNA-seq data were processed using the Cellranger software package (v7.1.0). To capture lineage information, a custom reference genome was constructed by appending the 2171 lineage tracing cassette sequences to the mouse mm10 genome build. Both gene expression and lineage tracing cassette libraries were then aligned to this reference using default parameters. Since the reference contains all 2171 30-nt barcode sequences, the intBC of each lineage tracing cassette read can be determined based on which element in the reference it aligns to.

A custom Python script was used to extract lineage information from the BAM alignment file generated by Cellranger. This script iterates through all reads aligned to a tracing cassette sequence. For each read, the LM sequence at each ES was determined based on the alignment. For each unique molecular identifier (UMI), the number of supporting reads was calculated, and the LM for each ES was set to the LM with the most read support (ties broken randomly). UMIs were then aggregated by intBC, cellBC, and the set of LMs detected across the three ESs to generate a final LM table.

scRNA-seq LM quality control

To generate high-quality lineage trees from scRNA-seq data, rigorous LM quality control is essential due to common issues such as PCR errors, sequencing errors, and ambient RNA contamination in tracing cassette reads. Since the levels of erroneous reads can vary between experiments—depending on sequencing depth and cell quality—and between lineage cassette integrations (intBCs) due to differences in expression levels, we applied adaptive thresholding to filter alleles in the LM table. First, for each 10X capture, we fit a two-component Gaussian mixture model to the distribution of reads per UMIs per allele and removed alleles corresponding to the lower peak of the distribution. Next, for each 10X capture and intBC, we applied the mixture model to the distribution of UMIs per allele, again filtering out alleles in the lower peak. The first step eliminates PCR and sequencing errors, which are characterized

by low read support, while the second step removes alleles derived from ambient RNA, which exhibit low UMI support that varies according to the expression level of the intBC.

When a lineage tracer is read out using endogenous RNA expression, another source of error arises from the lag time between the integration of the LM into the genome and the point when RNA molecules containing that LM become the predominant species. This issue is exacerbated by the long RNA half-lives of LTCs, particularly in experiments conducted over short timescales, leading to errors in calling LMs that occur later in the experiment. To mitigate this problem, instead of resolving conflicting alleles by simply selecting the allele with the highest UMI count, we considered the edit state of the alleles. If the conflicting alleles differed at a single ES—where one allele was edited and the other was not—we selected the edited allele, provided it accounted for >20% of the total UMIs for that intBC in the cell.

With only 2171 distinct intBCs, duplicate integrations—where two tracing cassettes with the same intBC integrate into the genome of a single clone—can occasionally occur. If not filtered out, these duplicates could introduce errors, as they accrue different LMs, and we are unable to resolve which LM occurred at which integration. To address this, we identified and removed alleles from duplicate intBCs by detecting allele conflicts. For each clone and intBC, we calculated the total fraction of UMIs in cells containing more than one allele and excluded intBCs where this fraction exceeded 25%.

scRNA-seq cell quality control

Low-quality cells in scRNA-seq experiments were filtered out based on mitochondrial content and total UMI counts. Cells with >10% of UMIs assigned to mitochondrial transcripts were excluded. For the kinetics experiment, cells with <100 UMIs assigned to the transcriptome were also removed. The UMI threshold varied between experiments due to differences in transcriptome sequencing depth. For scRNA-seq of fully edited clones and lineage tracing cells with static barcodes, a threshold of 1000 UMIs was used, while a higher threshold of 5000 UMIs was applied to cells used in the in vitro heterogeneity analysis, reflecting their deeper sequencing coverage.

Lineage tracing enables robust and accurate detection of scRNA-seq doublets–droplets containing more than one cell–since conflicting alleles reveal the presence of multiple distinct lineages within a single droplet. Doublets were filtered out by identifying these conflicting alleles, as it is unlikely that two random cells would share the same set of LMs. Similar to the process used for removing duplicate intBCs, we calculated the total fraction of UMIs linked to intBCs with conflicting alleles for each cell and excluded cells where this fraction exceeded 25%.

epegRNA kinetics analysis

epegRNA libraries were processed using the "CRISPR Guide Capture" feature of Cellranger. epegRNA protospacer mismatch calls supported by <3 UMIs were filtered out and careful quality control performed to exclude cells with multiple epegRNA integrations as these cells could confound our estimation of protospacer mismatch kinetics. Cells with multiple epegRNA calls were removed as well as cells with lineage cassette reads where more than one ES had an LM installed.

To calculate the edit fraction over time for each epegRNA protospacer mismatch in each cell line, we first determined the fraction of intBCs with an edit in each cell. We then calculated the mean edit fraction across cells for a given epegRNA call at each time point. Time points with fewer than 20 cells for a particular protospacer mismatch were excluded due to excessive noise in these data points.

To estimate the edit rate for each epegRNA protospacer mismatch, we fitted a saturating exponential curve to the edit fraction data. The curve is defined by the equation:

$$f = s\left(1 - e^{-rt}\right)$$

where f is the edit fraction, t is the time point (days), r is the edit rate [edits per (ES * day)], and s is the saturation level which accounts for a small fraction of cells that do not undergo editing despite expressing the epegRNA. The curve was fitted using non-linear least squares, with s constrained to the range [0.8,1] and r constrained to the range $[10^{-4},0.6]$. Estimated edit rates in 4T1 and B16-F10 cells for each protospacer variant are reported in table S10.

Generation of 4T1 PEtracer cells with high numbers of integrated tracing cassettes

4T1 cells were maintained in typical culture conditions. Infections with lineage tracing cassette lentiviral libraries were carried out by plating 100,000 cells per well of a six-well dish 18 to 24 hours prior to transduction. Immediately before adding the virus, media was changed on these cells to include 10ug/ml polybrene (Millipore Sigma). Fifty microliters of high titer virus was incubated on cells for 24 hours prior to a media change and outgrowth of cells. The brightest 10% of mCherry expressing cells were sorted 1 week after transduction to isolate cells with high integration numbers. Representative sorting gates are shown in fig. S16. This process was repeated serially to increase integration count until the average number of integrations in the population was over 10 as evaluated by qPCR using the following primer pairs and a single multiplicity of infection (MOI) line where infection rate was ~20% as an internal 1x integration control. qPCR reactions were carried out using 50 ng of genomic DNA isolated using DNeasy blood and tissue kit (Qiagen) with 500 nM primer mix and 1X DyNAmo ColorFlash SYBR Green Master Mix (Thermo) and the following cycling conditions: 95°C for 7 min, [95°C for 10 s, 60°C for 30 s]x40. Relative integration number was calculated by the ddCt method compared to β-actin.

Mouse β-Actin Control Forward: TTTGATGTCACGCACGATTT
Mouse β-Actin Control Reverse: AGGGCTATGCTCTCCCTCAC
mCherry Forward: AGGGCGAGATCAAGCAGAG
mCherry Reverse: CTCGTTGTGGGAGGTGATGT

Generation of 4T1 PEtracer cells edited to completion (fully edited cells)

4T1 were nucleofected to introduce a no mismatch 24-mer pegArray (all epegRNAs in the array had protospacers with no mismatches that would slow their editing rates). Nucleofections with 4T1 cells were carried out using the Lonza 4-D nucleofector according to manufacturer's instructions with the SE buffer, pulse code CM-150, 1.8 μg of donor pegArray plasmid and 180 ng of PiggyBac transposase plasmid (SBI). Cells were allowed to recover for 48 hours, split for >10 days to ensure loss of unintegrated plasmid, and BFP+ cells were isolated using FACS on a BD FACS Aria II (BD Biosciences). BFP+ cells were then engineered with high numbers of integrated tracing cassettes (LTCv2) by serial lentiviral infection as described above ("Generation of 4T1 PEtracer cells with high numbers of integrated tracing cassettes"). BFP+/mCherry+ cells were plated and infected with PEmax-P2A-GFP lentivirus as described for the tracing cassette lentivirus. When infecting cells with PEmax lentivirus, we aimed for 20-30% infection rates in these cells, corresponding to a single integrant per cell for the large majority of PE-containing cells. BFP+/mCherry+/GFP+ triple positive cells were FACS sorted >5 days posttransduction, passaged for >3 weeks, and then re-sorted at 5 cells/well. Representative sorting gates shown in fig. S16. Clones were expanded, evaluated by bulk sequencing, and pooled to cover all unedited and edited states.

scRNA-seq of fully edited 4T1 cells

Fully-edited 4T1 cells were passaged in RPMI as described above. Prior to droplet-based single cell sequencing using the 10X Genomics Chromium Next GEM Single Cell 3′ Kit v3.1, cells were split, counted, and diluted according to manufacturer's protocols. Lineage tracing cassette libraries were generated as described above ("epegRNA mismatch library profiling to tune lineage tracing kinetics").

We anticipated recombination during cloning and lentiviral production steps and therefore registered a link between MERFISH and sequencing intBCs in fully-edited 4T1 cells by generating targeted sequencing libraries from amplified cDNA using T3 specific primers (located 5′ of both the MERFISH and sequencing intBCs).

Two successive rounds of PCR were used to prepare intBC libraries from amplified full-length cDNA. For the first PCR reaction (PCR1), 2 ul of amplified cDNA and 600 nM primer mix listed below (T3_10X_PCR1_1-4_F, 10X_PCR1_R) were used. Amplifications were performed using KAPA HiFi HotStart ReadyMix with the following cycle conditions: 95°C for 3 min, [98°C for 20s, 65°C for 15s, 72°C for 15s]x25, 72°C for 1 min. PCR2 amplification was carried out using 4 ul of a 1:25 dilution of PCR1 product and 300 nM of the reverse primer (10X_PCR2_R) listed below and Nextera i7 adapter PCR primers (listed in table S29) with the following conditions: 95°C for 3 min, [98°C for 20 s, 72°C for 30 s]x12, 72°C for 1 min.

scRNA-seq intBC amplification primers:

PCR1 forward (T3_10X_PCR1_1-4_F): GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG -[1N to 4N]-GTAATTAACCCTCACTAAAGG

PCR1 reverse (10X_PCR1_R): ACACTCTTTCCCTACACGACG

PCR2 forward (Nextera i7): CAAGCAGAAGACGGCATACGAGATNNNNNNNGTCTCGTGGGCTCGG

PCR2 reverse (10X_PCR2_R): AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC

Following PCR amplification, lineage tracing cassette and intBC libraries were purified and size-selected using SPRI magnetic beads (0.9x single-sided selection) and quantified by BioAnalyzer (Agilent) to assess the size and purity of final libraries and Qubit double-stranded DNA high sensitivity Kit (Thermo Fisher Scientific) to determine the concentrations. Libraries were diluted to 4 nM and sequenced on a Nextseq 2000 (Illumina). All PCRs were first monitored to determine optimal cycle number by qPCR by adding 0.6x SYBR Green (Thermo) to the PCR reactions.

Reads from the T3 library were processed similarly to reads from the standard tracing cassette library, with read per UMI and UMI thresholds determined using Gaussian mixture models. The 183-nt MERFISH barcode for each T3 read was identified based on alignment, while the 30-nt sequencing barcode was extracted from the read sequence and corrected using a whitelist of known barcodes.

To map 183-nt MERFISH barcodes to 30-nt sequencing barcodes, we assigned each MERFISH barcode to a sequencing barcode in each cell based on UMI support. We then calculated the fraction of cells supporting each mapping for each clone. This approach allowed us to confidently assign MERFISH barcodes to 29 of the 31 sequencing barcodes present in the fully edited 4T1 cells, revealing multiple instances of barcode swapping (fig. S5B and table S11). In cases of barcode swapping, integrations are referred to by their MERFISH barcode identifier.

Clone calling using intBCs

Non-negative matrix factorization (NMF) was used to call clones which we define as groups of phylogenetically related cells that share a set of intBCs. Given the UMI counts for each intBC in each cell after quality control filtering, we represented the data as a cell-by-intBC matrix X. NMF was then applied to factorize X into the product of two matrices, W and H:

$$X = WH$$

Here, W is a cell-by-clone matrix representing the assignment of cells to clones, while H is a clone-by-intBC matrix representing the assignment of intBCs to clones. The number of clones for each experiment was manually determined based on visual inspection of the matrix X.

To filter out scRNA-seq doublets, which often contain cells from distinct clones, we identified cells with intBC sets that more closely matched the combination of two clones than any single clone. First, cells were putatively assigned to clones using the matrix W. Next, for each intBC, we assigned it to a clone if it was detected in >50% of the cells within that clone. Using this clone-to-intBC mapping, we calculated the Jaccard similarity between the set of intBCs detected in each cell and the intBC sets for both individual clones and all pairwise combinations of clones. Cells were assigned to the clone with the highest Jaccard similarity unless the highest match was to a pair of clones, in which case the cell was marked as a doublet. If the Jaccard similarity between a cell of interest and all possible clone assignments was <0.5, the cell was marked as unassigned.

Quantification of lineage tracing data accuracy in fully edited cells

To quantify the accuracy of lineage tracing data in fully edited 4T1 cells, we generated a whitelist (table S11) for each clone containing the LMs installed at each ES for each intBC. This whitelist was derived from the scRNA-seq data, where cells were grouped into four clones based on their intBCs, as described above. For each clone, we then identified the most common LM at each ES for each intBC, which formed the basis of the whitelist.

We assessed the accuracy and sensitivity of both intBC and LM detection using this whitelist. Based on each cell's clone assignment, we identified the expected set of intBCs and LMs and compared these to the actual intBCs and LMs detected. For intBCs, we calculated the true positive and false negative rates from the fraction of expected intBCs that were detected, and the false positive rate from the fraction of detected intBCs that were not expected. For LMs, we calculated overall call accuracy, as well as accuracy per ES and per LM. In scRNA-seq data, the LM detection rate matches the intBC detection rate, as LMs can be called for all detected intBCs. However, in imaging data, the LM detection rate is slightly lower than the intBC detection rate, since for some amplicons the intBC can be accurately called but the LM at one or more ESs cannot.

Microscope setup for image acquisition

Image acquisition was performed using a custom-built microscope system, as previously described (*151*) with some modifications. The system was built around a Nikon Ti2-U microscope body with a Nikon CFI Plan Apochromat Lambda D 60x oil immersion objective with 1.42 NA. The system also included a Nikon CFI Plan Fluor 10x objective with 0.3 NA for quick identification of the sample position. Illumination of samples was through a Lumencor CELESTA light engine (a fiber-coupled solid-state laser-based illumination system) with the following wavelengths: 405, 477, 545, 637, and 748 nm. This system was used with a penta-bandpass dichroic mirror (Semrock, FF421/491/567/659/776-Di01-25X36), a penta-bandpass emission filter (Semrock, FF01-441/511/593/684/817-25), and a penta-bandpass excitation filter (Semrock, FF01-391/477/549/636/741-25). A scientific CMOS camera (Hamamatsu C14440 with factory calibration for single-molecule imaging) was used for image acquisition. Each camera field of view (FOV) consisted of 2304 by 2304 pixels, with a camera pixel corresponding to 107 nm in the X and Y dimensions in the imaging plane for the 60x oil immersion objective with 1.42 NA. Sample position in three dimensions was controlled using a XYZ stage (Ludl Electronic Products). A custom-built auto-focus system was used to maintain a constant focal plane over prolonged time intervals. This was achieved by comparing the relative position of two IR laser (Thorlabs, LP980-SF15) beams reflected from the glass-fluid interface and imaged on a separate CMOS camera (Thorlabs, CS165MU1).

These different components were controlled using a National Instruments Data Acquisition card (NI PCIe-6321, X Series DAQ) and custom software (see the "Software for controlling experimental components" section below).

MERFISH fluidics system configuration

The fluidics system consisted of several core components similar to previous designs with some modifications (*151*): a syringe pump to

control flow speed and volume; a set of daisy-chain-connected valves to select flow-in buffers; a flow chamber in which the sample was mounted; a set of tubing, connectors, and needles; and a 3D printed chassis to hold a 24-well deep-well plate (Agilent, 204023-100). Four MVP4 High Torque valves (Hamilton, 94747-02) with eight-way connector heads (Hamilton, 97472-01) were daisy-chain connected (port 8 of each valve is connected to the input of the next valve), and the output of the first valve was connected to the FCS2 flow chamber (Bioptechs, 03060319-2-NH-30) in which samples were placed together with a 0.5-mm-thick rubber gasket (Bioptechs, DIE# F18524). The chamber output was connected to the input port of the syringe pump (Hamilton, PSD6) which controlled the flow speed and volume. The output port of the syringe pump was further connected to a waste collecting vessel.

This system was controlled using a custom software (see the "Software for controlling experimental components" section below). Overall, this system allowed for 24 rounds of hybridization. We also added two three-way adaptor sets (Bioptechs, 162003-1) and connected these two adaptors directly with tubing to allow fluidic bypass of the FCS2 chamber, which allows us to perform washing of the fluidic system without unmounting the sample if needed.

Software for controlling experimental components

All system components were controlled using custom-built software available at https://github.com/ZhuangLab/storm-control. The software package is composed of several main modules that work in concert:

- "Hal" is the software package used to control and synchronize all illumination and microscope components. We note that in some cases it is necessary to install drivers for components which are not included in this package, for example, the camera for autofocus system (Thorlabs, CS165MU1) requires Thorlabs official drivers. Specific imaging parameters including laser intensities, z-scan steps, exposure time for each frame and the sequence of shutter wrapped in .xml files are loaded and executed by Hal. These parameter files were generated by Python scripts provided in our Github repository (see "Code availability" below).
- "Steve" is a module used to generate mosaic images (i.e., a composite of many individual FOVs) of a sample to select FOV for imaging in experiments.
- "Kilroy" is the software used to control fluidics components, and to define pre-programmed sequences of operations to be performed as sets (e.g., the set of operations that initiate a new round of hybridization).
- "Dave" can issue commands to both Hal and Kilroy and is used to automate the performance of data collection by preprogramming the complete set of fluidics system and microscope operations used for an experiment (i.e., the order and time-lag in which fluidics and imaging steps are to be performed).

The general outline of an experiment is as follows: before the experiment starts, Hal and Kilroy are loaded with the parameters and specifications to be used. After the sample is loaded and the chamber is filled with the imaging buffer, a mosaic image of the DAPI channel is taken using Steve, and FOVs of interest are selected. A file is then generated to specify the sequence of operations throughout the entire experiment and is loaded to Dave, together with the coordinates of the selected regions of interest. The rest of the experiment is automated. If the number of rounds in an experiment exceeds the capacity of the flow system, we can specify a Dave file to automatically run up to the capacity of the system, then use the bypass tubing to clean up the fluidic system without perturbing the mounted sample. After cleaning up, all buffers are replaced and a new Dave file is created. This is repeated until all rounds of imaging are completed.

Design of intBC encoding probes

A combinatorial detection scheme for our 2171 intBCs was inspired by MERFISH experiments (*58*) and used a modified Hamming distance 4, Hamming weight 6 21-bit codebook derived from covering designs (https://www.dmgordon.org/cover/, with parameters: v=21, k=6, t=5). Each of the 21 total bits in the codebook are assigned to a distinct readout sequence and each intBC was assigned to a six-bit code in the codebook (table S6). We designed six encoding probes per intBC consisting of a 30-nt targeting sequence and three readout sequences. For each intBC, all assigned readout sequences are represented across the set of encoding probes. The 183-nt MERFISH barcode within each integration barcode was partitioned into six non-overlapping 30-nt segments. These 30-nt sequences have been tested for cross-hybridization during intBC design procedures, therefore the reverse-complement of these 30-nt sequences were directly used as targeting sequences in probes. Readout sequences were appended to each targeting sequence separated by an "A," with all assigned readout sequences evenly represented across the set of encoding probes designed for that intBC. Twenty-nucleotide forward and reverse primers were then added to the 5′ and 3′ of these probes. Then the fully assembled probes were further screened for cross-hybridization, so that any probes with a 17-nt segment reverse-complementing to other probes were removed. Full probe sequences can be found in table S12.

Design of LM and unedited probes

A distinct probe was designed to hybridize to each LM installed at its cognate ES. Specifically, ES sequences modified with each LM were generated in silico and probes were designed to hybridize to each modified ES. Probes were designed to be 20 nt in length and centered around each 5-nt inserted LM. For each LM, a specific MERFISH readout sequence was chosen and two copies of this readout sequence were appended to the 5′ side of each ES/LM probe. All LM probes were directly purchased from IDT. Several combinations were tested and the final set of probes with the highest signal to noise ratios (SNRs) were merged. Full probe sequences can be found in table S13.

Design of tracing cassette common sequence probes

Target sequences of common bits were directly chosen from the 500-nt common sequence located 3′ of the ESs in the LTC (Addgene #238548). Common sequence probes were designed to target 24 30-nt sequences within this 500-nt common sequence. Two 20-nt MERFISH readout sequences (table S16) were appended to the 5′ and 3′ of these targeting sequences flanked by a set of forward and reverse primers. Full probe sequences can be found in table S14.

Encoding probe amplification

Encoding probes were ordered as oligo pools from Twist Bioscience. Each oligo pool was first dissolved in 100 ul TE buffer (Thermo Fisher Scientific, AM9849). Encoding probes were amplified from the template library using a previously described amplification protocol (*151*) with additional modifications:

1) The initial oligo pool was amplified using limited-cycle PCR for approximately 11 to 15 cycles dependent on the library size, where the reverse primer used for this amplification included a T7 promoter sequence to enable in vitro transcription in the next step. PCR primer sequences are listed below.

2) The resulting PCR product was purified via column purification (DNA Clean & Concentrator-25, Zymo Research, D4033).

3) The purified PCR product underwent further amplification and conversion to RNA by a high-yield in vitro T7-mediated transcription reaction (HiScribe T7 Quick High Yield RNA Synthesis Kit, NEB, E2050).

4) The resulting RNA product was purified by column (Monarch RNA Cleanup Kit, NEB, T2050).

5) The purified RNA product was converted back to single-stranded DNA by a reverse transcription reaction (Maxima H Minus Reverse Transcriptase, Thermo Fisher Scientific, EP0753) using the forward primer from step 1 in this protocol. The final product was purified by adding 3x volume of the self-made SPRI beads, eluted in TE buffer

followed by an ethanol precipitation to concentrate to 25 nM for each probe in TE buffer. This probe stock could be stored in −20°C. All primers were purchased from IDT.

Probes for common sequence in lineage cassette were amplified by:

Forward primer: CGGGTTTCGTTGCGCACACC

Reverse primer: TAATACGACTCACTATAGGGCTTGTGCATCGCGCCAAAGA

Probes for intBC were amplified by:

Forward primer: CCCGCAATGGCTGACAACCG

Reverse primer: TAATACGACTCACTATAGGGATTGCCGCATGGTTTCCG

Probes for 124-gene MERFISH library were amplified by:

Forward primer: CCCGCAATGGCTGACAACCG

Reverse primer: TAATACGACTCACTATAGGGATTGCCGCATGGTTTCCG

Probes for 175-gene MERFISH library were amplified by two sub-pools:

Amplification of 150-gene Sub-pool Forward primer: CCCGCAATGGCTGACAACCG

Amplification of 150-gene Sub-pool Reverse primer: TAATACGACTCACTATAGGGATTGCCGCATGGTTTCCG

Amplification of 25-gene Sub-pool Forward primer: CGGGTTTCGTTGCGCACACC

Amplification of 25-gene Sub-pool Reverse primer: TAATACGACTCACTATAGGGCTTGTGCATCGCGCCAAAGA

Glass coverslip cleaning and silanization

40-mm-diameter, 0.17-mm-thickness glass coverslips (Bioptechs, 40-1313-03192) were first cleaned by placing coverslips in a plastic holder (Entergris, A23-0215) and performing 3x 5-min rinses with MilliQ water in a glass dish (Electron Microscopy Sciences, 70312-31) prior to a 30-min incubation in a chemical hood at room temperature with a mixture of equal volumes of Methanol (Sigma, 34860) and 37% HCl (Sigma-Aldrich 258148). After this incubation, 4x 5-min MilliQ water rinses were performed prior to 2x 5-min washes with 70% ethanol. Coverslips were gently blown dry using compressed air and left to fully dry.

Cleaned coverslips were placed in a inert plastic holder and incubated for 30 min at room temperature in a mixture of 1 liter chloroform (Sigma-Aldrich 319988), 1 ml triethylamine (Sigma-Aldrich 81101), and 2 ml allyl-trichlorosilane (Sigma-Aldrich 107778) which were all stored and handled according to provider's instructions. Using glass pipetting agents is essential to generating high-quality coverslips. Following this incubation, coverslips were washed by performing 2x 5-min chloroform washes followed by a 1x 5-min wash in 100% ethanol. After this ethanol wash, coverslips were gently blown dry by compressed air and left to fully dry. Following silanization, coverslips were stored in a dry, dark place in a sealed container with a dedicated set of desiccant cartridges (VWR, 76538-672).

Poly-D lysine coating silanized glass coverslips

Silanized coverslips were incubated with 0.1 mg/ml Poly-D lysine (PDL; Thermo Fisher Scientific, A3890401) overnight at room temperature in a 3D-printed box with parafilm sealing top. Following PDL coating, coverslips were cleaned with 3x 5-min rinses with MilliQ water followed by a 1x rinse with 100% ethanol before being gently blown dry with compressed air and left to dry. Coverslips were stored in 3D-printed cases at 4°C with the same desiccant cartridge and sterilized with ultraviolet light for 20 min immediately prior to use.

Seeding of fully edited cells prior to imaging

Fully edited cells were cultured as described above prior to seeding 400,000 cells on cleaned, silanized, and PDL-coated 40-mm-diameter round coverslips in 60-mm-diameter petri dishes (Corning, 430166) in RPMI media supplemented with 10% FBS and 0.1% PSG. Samples were then prepared for FISH-based lineage readout using either the "Zombie" or "In-gel T7" protocols described in the following sections. Fifty-nine FOVs were imaged for one coverslip prepared with the "Zombie" protocol and 69 FOVs for one coverslip prepared with the "In-gel T7" protocol as enabling comparison of these methods.

Sample preparation for Zombie experiments

The Zombie protocol was adapted from previously published methods (*42*, *56*). In brief, samples were washed 2x with 1x PBS for 5 min each and subsequently fixed with a 3:1 (vol:vol) mix of methanol and acetic acid at room temperature for 20 min. Cells were then washed 4x with 1x PBS prior to a 1x wash with nuclease-free water. Samples were subsequently incubated with T7 in vitro transcription mix (MEGAscript Transcription Kit; Invitrogen) supplemented with 2.5 mM aminoallyl-UTP (Thermo Fisher Scientific, R1091) at 37°C for 16 hours. After transcription, cells were fixed with 4% formaldehyde in PBS for 20 minutes at room temperature followed by 2x washes with 1x PBS prior to the same hybridization as in the section "Lineage probe hybridization" below.

Sample preparation for in-gel T7 amplification

Samples were washed 2x with 1x PBS for 5 min each and then fixed by treatment with 4% paraformaldehyde in 1x PBS for 10 min at room temperature. Samples were subsequently washed three times with 1x PBS and stored in 1x PBS supplemented with 1:500 RNase Inhibitor Murine (NEB, M0314L) for up to 8 weeks.

The fully edited cells fixed on coverslip were embedded in a hydrogel prior to tissue clearing to reduce tissue autofluorescence and off-target probe binding. First, the sample was incubated in gel solution consisting of 300 mM NaCl, 4% (v/v) of 19:1 acrylamide/bisacrylamide, 60 mM Tris-HCl pH 8, 0.1% (v/v) TEMED, 1/40,000 (v/v) 505/515-nm fiducial beads (Invitrogen F8803), and 0.5% (w/v) ammonium persulfate for 5 min at room temperature. We then placed a 60 ul droplet of the gel solution onto a hydrophobic glass slide treated with GelSlick (Lonza). The coverslip bearing the sample was then inverted onto the droplet to form a uniform layer of monomer solution. The sample was allowed to polymerize completely for >1 hour at room temperature.

The coverslip bearing the gel-embedded sample was then removed from the glass slide with a thin razor blade. The sample was then incubated in digestion buffer containing 2% (w/v) SDS, 0.5% (v/v) Triton-X-100 (Sigma, X100), and 1% (v/v) Proteinase K (NEB P8107S) in 2x SSC for 24 hours at 47°C and a subsequent 24 hours at 37°C. Following digestion, the sample was washed by performing 6x 20-min washes with 2x SSC buffer supplemented with 0.2% (v/v) Rnase Inhibitor Murine (NEB) at room temperature. The sample was further digested in 800 mM Guanidine HCl (IBI Scientific, IB05080), 1 mM EDTA, 50 mM Tris-HCl pH=8.0, and 0.5% Triton-X-100 supplemented by 1% (v/v) Proteinase K for 24 hours at 37°C.

After Guanidine HCl digestion, samples were transferred to new petri dishes with fresh 2x SSC for one quick wash followed by a 10-min incubation in 2x SSC supplemented with 1 mM phenylmethylsulfonyl fluoride (PMSF) protease inhibitor (Thermo Fisher Scientific, 36978). Samples were then thoroughly washed by performing 5x 15-min rinses with 2x SSC on a shaker followed by 3x 15-min washes with nuclease free water on the shaker. The remaining water was aspirated from the coverslips, and each coverslip was inverted onto a 100-ul droplet of T7 in vitro transcription mix (MEGAscript Transcription Kit; Invitrogen AM1334) supplemented with 2.5 mM aminoallyl-UTP (Thermo Fisher Scientific, R1091) on a parafilm-covered 60-mm-diameter petri dish. Transcription was carried out for 16 hours at 37°C. Optionally, samples were remounted onto fresh T7 transcription mixture after this 16-hour incubation for an additional 3 hours at 37°C.

After transcription, samples were fixed with 4% formaldehyde solution in 1x PBS for 20 min at room temperature followed by 2x washes with 1x PBS. After these washes, probe hybridization was carried out as described in the section "Lineage probe hybridization" below.

Lineage probe hybridization

Following either the Zombie or in-gel T7 amplification protocols above, probes for lineage intBCs, common sequences, and LMs were hybridized similar to published MERFISH protocols (*152*). Two separate hybridizations were performed: in the first, probes for the intBCs and common sequences were hybridized, while in the second LM-encoding probes were hybridized.

For the first hybridization, samples were washed once with 2x SSC prior to equilibration in encoding-probe wash buffer [30% (v/v) formamide (Thermo Scientific, AM9344) in 2x SSC] supplemented with 0.1% Tween-20 (Sigma, P9416) for 10 to 15 min at room temperature. Wash buffer was then aspirated from the coverslip, and coverslips were inverted with the sample side facing down onto a 50-ul droplet of probe mixture on a parafilm-covered 60-mm petri dish. The probe mixture comprised encoding probes for each intBCs at 1 nM and encoding probes for the two-color common sequence at 10 nM in 2x SSC with 30% v/v formamide, 0.1% w/v yeast tRNA (Life Technologies) and 10% v/v dextran sulfate (Sigma, D8906). Probe hybridization was carried out by incubating at 37°C for 24 hours in a humidified chamber. After hybridization, samples were washed 1x with encoding-probe wash buffer for 30 min at 47°C to remove excess probes prior to a 1x wash with 2x SSC before proceeding directly to LM probe hybridization.

Samples were equilibrated in LM-probe wash buffer [10% (v/v) formamide in 2x SSC] supplemented with 0.1% Tween-20 for 10 to 15 min at room temperature. Wash buffer was then aspirated from the coverslip, and each coverslip was inverted with the sample side facing down onto a 50-ul droplet of probe mixture on a parafilm-covered 60-mm petri dish as above. This probe mixture comprised 300 nM of each LM probe (directly purchased from IDT) in 2x SSC with 30% v/v formamide, 0.1% w/v yeast tRNA (Life Technologies) and 10% v/v dextran sulfate (D8906, Sigma). Samples were incubated at 37°C for 24 hours in a humidified chamber. After hybridization, the sample was washed 2x with the encoding-probe wash buffer for 15 min at room temperature to remove excess probes.

Following these two hybridizations, samples were embedded in a second 4% poly-acrylamide gel as described in the section "In-gel T7 amplification for fully- edited cells" above to further immobilize T7 amplicons. After this embedding, samples can be stored at 4°C in 2x SSC supplemented with 1:1000 Rnase Inhibitor Murine for up to 4 days.

Imaging of integration barcodes and lineage marks

After lineage probe hybridization and secondary gel-embedding, samples were assembled into the FCS2 flow chamber (Bioptechs, 060319-2). Samples were imaged by the home-built imaging platform described in the "Microscope setup for image acquisition" and "MERFISH fluidics system configuration" sections. All fluid exchanges on the microscope use the custom-built fluidics system described in the "Fluidics system configuration" section. For each coverslip, nuclear staining by DAPI was used to help select FOVs of interest from mosaic images generated at 10x magnification using "Steve" as described in "Software for controlling experimental components".

Throughout the course of imaging, we performed multiple rounds of hybridization and imaging. Specifically, we imaged 2 colocalizing bits for common sequences used to nominate potential amplions in round 0; 21-bit combinatorial scheme for integration barcode readout in rounds 1 to 7, and 27 sequential bits for LMs and unedited states in rounds 8 to 16 (three colors per round). We designed 51 adaptor probes to convert bit-specific readout sequences in the encoding probes into common readout sequences. Each adaptor probe consists of: a 5′ sequence that is complementary to one of the bit-specific readout sequences used in encoding probes, and a 3′ sequence with two copies of the reverse-complement of one common readout sequence (table S16) (*151*). We designed three common readout sequences in total. Each common readout probe contains a readout sequence that is complementary to the adaptor probe and has a fluorescent dye (Alexa750, Alexa647 or Atto565) linked to the 5′ of the oligo via a disulfide bond (table S15).

For each round, samples were imaged as follows unless otherwise described:

1) Samples were hybridized with three adaptors probes labeling three different bits at 100 nM in secondary hybridization buffer (2x SSC, 30% v/v formamide) for 15 min at room temperature.

2) Samples were washed with the secondary hybridization buffer for 3 min at room temperature.

3) Samples were then hybridized with a set of fluorescently labeled oligonucleotide probes termed "common readout probes" (synthesized by IDT) (table S15) at 20 to 25 nM in secondary hybridization buffer for 15 min at room temperature.

4) Samples were washed again with the secondary hybridization buffer, for 3 min at room temperature.

5) Samples were then incubated and kept in imaging buffer comprising 5 mM 3,4-dihydroxybenzoic acid (Sigma, P5630), 100 μM of Trolox and Trolox quinone mixture (generated by UV radiation of a Trolox solution ~50 μM of Trolox quinone final quantified by UV-vis spectrometry; Sigma, 238813), 1:500 recombinant protocatechuate 3,4-dioxygenase (rPCO; OYC Americas, 46852904), 1:1000 Murine RNase inhibitor (NEB, M0314L), 50 mM Tris-HCl pH=7.5 (Thermo Fisher Scientific, 15568025), and 5 mM NaOH (to adjust pH to 8.0; Sigma, S2770) in 2x SSC.

6) Samples were imaged using a 60x oil immersion objective, and signals for the common readout probes and fiducial beads were collected in the 748-, 637-, 545-, and 488-nm channels, respectively at a rate of 10 Hz. All channels were imaged for 50 to 60 consecutive 0.6-μm-thick z-stacks for each FOV of interest.

7) After imaging all FOVs, the signal from the previous round was extinguished by incubating the sample in cleavage buffer comprising 2x SSC, 30% formamide supplemented with 50 mM Tris (2-carboxyethyl) phosphine (TCEP; Sigma, 646547) for 10 min at room temperature. TCEP treatment cleaves the disulfide bond connecting fluorophores to readout probes. Cleavage buffer also contained 167 nM common readout probes without fluorophores to block unoccupied common readout sequences on the adaptor probes from interfering with the next round of hybridization.

All rounds of imaging were carried out as described above except for the first round. For the first round, before the imaging buffer step (step 5), samples were additionally washed in 2x SSC once and then incubated in 2x SSC containing 2.5 μg/ml DAPI (Thermo Fisher Scientific, D1306) for 10 min to stain nuclei. DAPI signal was collected in the 405-nm channel, and either the 405-nm channel or all channels in this round were imaged for 50 to 60 consecutive 0.6-μm-thick z-stacks for each FOV of interest, depending on the experiment. Nuclear DAPI signal was imaged together with common sequences in round 0, which was later used as reference for segmentation and fine alignments described here: "Sample alignment between MERFISH and lineage imaging."

Nuclei segmentation for in vitro samples

Cellpose (v3.1.0) was used to segment cell nuclei in 3D format based on DAPI signal (*98*). For each FOV the DAPI image was adjusted to account for the illumination profile of the microscope and then downsampled by a factor of four in the x-y plane. Cellpose was run on a GPU using the "nuclei" model with a diameter of 25 pixels and a cell probability threshold of −4. The resulting nuclei masks were converted into polygons with a tolerance of 0.5 pixels.

Nuclei polygons from all FOVs were then aggregated. Since adjacent FOVs overlap, the same nuclei could be segmented in multiple FOVs, leading to overlapping polygons that required correction. For large overlaps (polygons from neighboring z planes that overlap >20% of their area), the two polygons were merged, as they were likely the same nuclei segmented twice. For smaller overlaps, the intersecting volume was subtracted from the smaller polygon, as these were likely segmentation

errors. After aggregating all polygons and correcting overlaps, polygons with a volume <500 μm^3 were removed.

3D nuclei polygons, represented as a collection of 2D polygons for each z-slice, were used to assign T7 amplicons and MERFISH transcripts to cells, as described in detail below. For 2D visualization, nuclei outlines were generated by computing the union of all 2D polygons along the z-axis.

T7 amplicon detection and quantification

After T7 amplification, each LTC integration should be marked by a local population of RNA molecules centered at the site of transcription. These amplicons were detected using hybridization probes targeting the common sequence shared by all integrations. To maximize the detection rate, the common sequence was probed in two independent rounds of hybridization. To identify amplicons, the two hybridization images were maximum-projected, and a 2D unsharp mask (sigma = 10 pixels) was applied to reduce background. Spherical intensity peaks were then identified in 3D format using the Laplacian of Gaussian method (sigma = 2 to 10 pixels, minimum intensity = 200). The spot detection parameters were optimized to prioritize recall of real amplicons, accepting false positives that could be filtered out during integration barcode decoding.

Following amplicon detection, the intensity of each hybridization round for each amplicon was quantified. A 2D unsharp mask (sigma = 10 pixels) was applied to each channel, and then the maximum signal intensity within a 10-pixel diameter circle in the x-y plane centered at each amplicon position was calculated. This quantification was performed in 3D format to resolve amplicons that overlapped in the z-dimension. To correct for x-y drift between imaging rounds, fiducial beads imaged in each round were used for registration. The drift was calculated using phase cross-correlation between bead images in each round compared to the first round, and amplicon positions were adjusted accordingly.

T7 amplicon integration barcode decoding

The intBC for each T7 amplicon was determined by probing the 183-nt MERFISH barcode across three rounds of three-color imaging. After detecting potential T7 amplicons and quantifying their intensity across the 21 bits as described above, data from all fields of view were combined into a *N-by-21* intensity matrix, X, where N is the total number of potential amplicons. With this representation, intBC decoding becomes an optimization problem, where we seek to assign each row in X to the most likely intBC codeword (21-bit Hamming distance 4, Hamming weight 6 code) accounting for variability in T7 amplification and probe hybridization efficiency across rounds. Following the approach of Moffitt *et al.* (2016) we use an expectation maximization (EM) algorithm to solve this optimization problem (*153*).

First, X is color normalized to correct for differences in laser power and fluorophore intensity by rescaling each column so that its mean intensity is consistent across channels. Then, to account for variability in amplicon intensity and probe hybridization efficiency, X is normalized by the dot product of mean spot and bit intensities:

$$X_{corrrected} = X \,/\, \left(s_{\text{intensity}} \bullet b_{\text{intensity}}\right)$$

Here, $s_{intensity}$ is defined as the mean intensity of each spot across the "1" bits indicated by its corresponding codeword, and $b_{intensity}$ is the mean intensity of each bit across spots (after normalization by spot intensity). Initial estimates are obtained by setting $s_{intensity}$ to the 95th percentile of each row and $b_{intensity}$ to the 95th percentile of each column (after scaling by $s_{intensity}$). Finally, $X_{corrected}$ is adjusted for each bit's noise profile:

$$X_{corrrected} = \left(X_{corrected} - 1/b_{snr}\right) / \left(1 - 1/b_{snr}\right)$$

where b_{snr} is the signal-to-noise ratio for each bit; values in $X_{corrected}$ >1 are clipped to 1.

In the expectation step, $X_{corrected}$ is used to assign each spot to an intBC codeword from the predefined list of 2171 valid binary codewords (listed in table S6). A KDTree is used to efficiently perform a nearest-neighbor search that identifies the intBC codeword minimizing the Euclidean distance to each column in $X_{corrected}$. Spots with a distance ≥1.7 between the codeword and corrected intensity vector are masked out. Spots with a distance ≥1.7 are masked out, a threshold that balances error correction with the exclusion signals that do not correspond to true integration amplicons.

In the maximization step, intBC identity assignment in the previous expectation step is used to update $s_{intensity}$, $b_{intensity}$, and b_{snr}, where $s_{intensity}$ is recalculated for all rows, while only spots with confidently assigned intBCs contribute to the updates of $b_{intensity}$ and b_{snr}. For each bit, b_{snr} is computed as the ratio of the mean intensity for spots where the bit is present ("1" in the codeword) to the mean intensity for spots where it is absent ("0" in the codeword).

After 10 iterations—sufficient for convergence—each amplicon is decoded based on its assigned intBC. Amplicons with a matching distance of ≥1.7 are filtered out and excluded from downstream analysis.

T7 amplicon LM decoding

Each LM is detected by a separate hybridization probe, so decoding the eight LMs for each ES as well as the unedited state requires nine rounds of three-color FISH imaging (27 bits total, 9 bits per ESES × 3 ESES). For spots with confident intBC assignments, data from all fields of view were combined into an N-by-27 intensity matrix X, as described above. X is then color-balanced by rescaling each column so that its mean intensity is consistent across channels, correcting for differences in laser power and fluorophore brightness.

To select an LM for each ES, we extract the relevant columns from X to form an N-by-9 intensity matrix, X_{site}. Each row in X_{site} is normalized so that the sum of its values equals 1, accounting for variation in overall spot intensity. While simply choosing the LM corresponding to the maximum value in each row of X_{site} works well, we found that training a logistic regression classifier further improves decoding accuracy. The classifier learns to account for variations in probe affinity and cross-hybridization that affect the intensity profile of each LM across hybridization rounds.

We trained the logistic regression classifier using the known linkage between intBCs and LMs in our fully edited cells (see "Quantification of lineage tracing data accuracy in fully edited cells" for details). For each row in X_{site}, we determined the ground-truth LM based on the decoded intBC and then trained the classifier to predict this value. In vitro decoding accuracies for both our in-gel tissue-clearing protocol and the standard Zombie protocol were evaluated using fivefold cross-validation. The classifier was subsequently retrained on all in vitro data from the in-gel protocol and then applied to decode all subsequent experiments, including in vivo data from fully-edited cells, which serve as an out-of-sample control.

Separate classifiers were trained for each ES. Accuracies were reported for all decoded LMs as well as for the subset of LMs confidently decoded by each classifier (with prediction probability p ≥0.7). Only LMs with p ≥0.7 were used in downstream analyses such as tree reconstruction.

Assignment of T7 amplicons to cells

Decoded T7 amplicons were assigned to nuclei segmentation masks in 3D format using the geopandas package (v1.0.1). Using the 3D coordinates of each amplicon and the 3D polygons representing nuclei, amplicons falling within a polygon were assigned to the corresponding nucleus, then to account for segmentation errors, amplicons located just outside a mask (<4 μm) were assigned to the nearest nucleus. The Euclidean distance from each amplicon to its assigned mask was recorded, and amplicons outside segmentation masks were only used in

tree reconstructions if no amplicon with the same intBC was detected within the mask.

Imaging readout LM quality control

Similar to scRNA-seq LM quality control (see "scRNA-seq LM quality control" for reference), we identify a set of high-confidence LM calls for each intBC in every imaged cell. We first remove amplicons with poor LM detection by excluding those with a mean LM assignment probability <0.7 across the three ESs. To account for instances where the same amplicon is detected multiple times by the spot detection algorithm, we group intBCs with identical LM sets within each cell and retain only the brightest spot from each group. To eliminate duplicate integrations (i.e., two lineage tracing cassettes sharing the same intBC), we exclude intBCs detected in more than one amplicon with different LMs in >40% of the cells for a given clone. Additionally, to remove segmentation doublets—cases where two cells are not properly separated—we filter out cells with >50% conflicting amplicons. Finally, to uniquely assign LMs to intBCs, we resolve conflicting amplicons by selecting the brightest spot in each cell, with a preference for those contained within the nuclear mask.

After quality control the accuracy of image-based intBC and LM decoding was quantified using fully edited cells as described in the "Quantification of lineage tracing data accuracy in fully edited cells" section.

Cloning and evaluation of puromycin/blasticidin lentiviral static barcode libraries

Lentiviral plasmids expressing puromycin or blasticidin resistance genes were cloned with Gibson assembly as described above. We introduced a 10N random barcode that could be efficiently assayed using scRNA-seq-based capture of 3′ polyadenylated transcripts. These libraries were cloned by introducing oligonucleotide libraries ordered from IDT using Gibson assembly. The same strategy used for 5N epegRNA libraries described above (see section "Cloning of a comprehensive 1024 5-nt LM epegRNA libraries" for details) was used where top and bottom oligo libraries were ordered, cloned, and transformed before libraries were pooled, expanded through outgrowth for 8 hours and Midiprepped as detailed above. Lentiviral plasmid libraries were sequence confirmed as described for the epegRNA libraries with modest differences. As before, plasmid libraries were amplified with two successive rounds of PCR as done to sequence genomic loci. PCR1 was performed using Phusion Green Hot Start II High-Fidelity PCR 2x Master Mix with the following cycle conditions: 98°C for 2 min, [98°C for 10 s, 64°C for 20 s, 72°C for 1 min] where the number of cycles was empirically determined by qPR, followed by a 72°C for 2 min. PCR2, gel extraction, library preparation, and sequencing were carried out as before.

Puromycin 10N library top: CAGAAAGCCTGGCGCCTGACNNNNNNNNNGTTTAAACCCGCTGATCAGCCTCGACTG

Puromycin 10N library bottom: CAGTCGAGGCTGATCAGCGGGTTTAAACNNNNNNNNNNGTCAGGCGCCAGGCTTTCTG

Blasticidin 10N library top: TTATGTGTGGGAGGGCTAACNNNNNNNNNGTTTAAACCCGCTGATCAGCCTCGACTG

Blasticidin 10N library bottom: CAGTCGAGGCTGATCAGCGGGTTTAAACNNNNNNNNNNGTTAGCCCTCCCACACATAA

Forward primer for puromycin library amplification: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG -[1N to 4N]-CCTGGTGCATGACCAGAAAGC

Forward primer for blasticidin library amplification: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG -[1N to 4N]-GTGAATTGCTGCCCTCTGGTTATGTG

Reverse primer for puromycin and blasticidin library amplification: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG -[1N to 4N]-GCACAGTCGAGGCTGATCAG

To assess the balance of library elements, we used FastX toolkit (v0.0.14) to mask low quality bases (<q20) and collapse reads. We removed common flanking sequences and summed reads for unique barcode sequences. Lentivirus was produced from these plasmid libraries as described above ("Lentiviral production").

In vitro tracing with ground-truth barcoding readout with scRNA-seq

4T1 cells with high numbers of integrated LTCv1 tracing cassettes were nucleofected with a pegArray tuned to a 2- to 3-week timescale (24-mer comprises: ES 2 protospacer mutant 8, ES 3 protospacer mutant = 34, ES 1 protospacer mutant = 32; see table S10). >10 days postnucleofection, BFP+/mCherry+ double positive cells were isolated by FACS and expanded. On day 0 of the experiment, tracing cassette and pegArray containing cells were subsequently infected with PEmax-P2A-GFP lentivirus with an infection rate of ~30% and thus typically one integrant per cell as above. Forty-eight hours postinfection (on day 2), BFP+/mCherry+/GFP+ triple positive cells were sorted as two cells per well into 96-well tissue culture-treated plates (Corning). Representative sorting gates shown in fig. S16. On day 5, sorted cells were infected with 10 ul of the puromycin static barcode lentiviral library. Media was changed on Day 6 to prevent continued infection events. On day 7, cells were infected with 10 ul of the blasticidin static barcode lentiviral library. On day 8, cells were split into a fresh TC-treated 96-well plate to continue expansion of the clones. Robust wells were expanded until treatment on day 12 with 4 μg/ml blasticidin-containing selection media. Cells were passaged in blasticidin selection media until day 14 when cells were split into media containing both 4 μg/ml blasticidin and 2 μg/ml puromycin. Double selection continued until day 16, at which points cells were split, counted, and droplet-based single cell sequencing was performed by pooling one to three clones per capture for a total of six clones and three captures using the 10X Genomics Chromium Next GEM Single Cell 3′ Kit v3.1. Lineage tracing cassette libraries were generated as described above ("epegRNA mismatch library profiling to tune lineage tracing kinetics").

Sequencing and analysis of in vitro tracing with ground-truth barcoding readouts

Two successive rounds of PCR were used to prepare puromycin and blasticidin barcode libraries from amplified full-length cDNA. For the first PCR reaction (PCR1), 2 ul of amplified cDNA and 600 nM primer mix listed below (Puro_10X_PCR1_1-4_F/Blast_10X_PCR1_1-4_F, 10X_PCR1_R) were used. Amplifications were performed using KAPA HiFi HotStart ReadyMix with the following cycle conditions: 95°C for 3 min, [98°C for 20s, 65°C for 15s, 72°C for 15s]x18 to 20, 72°C for 1 min. PCR2 amplification was carried out using 4 ul of a 1:25 dilution of PCR1 product and 300 nM of the reverse primer (10X_PCR2_R) listed below and Nextera i7 adapter PCR primers (listed in table S29) with the following conditions: 95°C for 3 min, [98°C for 20s, 72°C for 30 s]x12, 72°C for 1 min.

intBC amplification primers:

PCR1 forward (Puro_10X_PCR1_1-4_F): GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG -[1N to 4N]-CCTGGTGCATGACCAGAAAGC

PCR1 forward (Blast_10X_PCR1_1-4_F): GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG -[1N to 4N]-GTGAATTGCTGCCCTCTGGTTATGTG

PCR1 reverse (10X_PCR1_R): ACACTCTTTCCCTACACGACG

PCR2 forward (Nextera i7): CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTCTCGTGGGCTCGG

PCR2 reverse (10X_PCR2_R): AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC

Following PCR amplification, LTC and Puro/Blast barcode libraries were purified and size-selected using SPRI magnetic beads (LTC libraries: 0.9x single-sided selection, Puro/Blast barcode libraries: 1.2x single-sided selection) and quantified by BioAnalyzer (Agilent) to assess the size and purity of final libraries and Qubit double-stranded DNA high sensitivity Kit (Thermo Fisher Scientific) to determine the concentrations. Libraries were diluted to 4 nM and sequenced on a Nextseq 2000 (Illumina). All

PCRs were first monitored to determine optimal cycle number by qPCR by adding 0.6x SYBR Green (Thermo) to the PCR reactions.

scRNA-seq processing and quality control were performed as described in the "scRNA-seq data processing" section, with additional puromycin and blasticidin barcode libraries aligned using Cellranger to a custom reference. This reference included puromycin and blasticidin resistance cassette sequences with "N"s indicating the positions of the 10-nt random barcodes. A custom Python script was used to extract the barcode sequence from each read aligned to these elements, generate a consensus barcode for each UMI, and aggregate UMIs by barcode for each cell. To separate barcode calls corresponding to real integrations from sequencing errors and ambient RNA, for each 10x capture a Gaussian mixture model was fitted to the distribution of UMIs per barcode per cell. Barcode calls corresponding to the lower peak were excluded as likely errors.

Quality control, clone calling, and phylogenetic tree reconstruction for each clone using neighbor joining were performed as described above.

Algorithm for calling barcode groups

A greedy algorithm was used to define barcode groups, which are clusters of cells sharing a set of puromycin or blasticidin barcode sequences. Starting with UMI counts for each barcode in each cell after quality control filtering, the algorithm iteratively performs the following steps until all cells are assigned to a barcode group or no viable seeds remain:

1) *Select a seed:* Choose the barcode detected in the highest fraction of unassigned cells that has not been used in a previous iteration.

2) *Identify the cluster:* Identify all other barcodes present in more than 80% of the cells where the seed barcode is detected.

3) *Assign cells:* Assign all unassigned cells to the barcode group if the barcodes in the cluster account for more than 80% of the total barcode UMIs in those cells.

4) *Validate the group:* Retain the barcode group if at least five cells are assigned; otherwise, discard the group and mark the cells as unassigned.

This algorithm accurately assigns the vast majority of cells to barcode groups, accounting for cases where groups are distinguished by multiple barcodes and instances where barcodes are occasionally reused between groups (fig. S6).

Quantification of agreement between phylogeny and barcode groups

The FMI, a metric for quantifying the similarity between two clusterings, was used to assess the agreement between the phylogeny and barcode groups derived from puromycin and blasticidin static barcoding. FMI is defined as:

$$FMI = \sqrt{\frac{TP}{(TP+FP)} \cdot \frac{TP}{(TP+FN)}}$$

where TP is the number of true positives, FP is the number of false positives, and FN is the number of false negatives.

To calculate the FMI score for each clone, barcode groups were compared to clades representing the descendants of the ancestral cell that originally received each barcode. The lowest common ancestor (LCA) of each barcode group was inferred by computing the FMI score for all internal nodes in the phylogeny and selecting the node that maximized the FMI for that group. For each clone, FMI scores were calculated separately for puromycin and blasticidin barcode groups, using phylogenies reconstructed with both neighbor joining (NJ) and UPGMA. Cells lacking a barcode excluded from the FMI calculation. FMI scores were also calculated after randomly permuting the barcode group associated with each cell in the phylogeny. FMI values are reported in table S17.

Inference of ancestral LM states

To estimate branch lengths in the phylogeny and collapse branches that are not supported by a LM, we inferred the ancestral LM state for each internal node in the tree using the Sankoff algorithm (*154*). This algorithm was chosen because it allows for low-probability LM-to-LM transitions, thereby reducing the impact of noise and errors in phylogeny reconstruction on ancestral state inference. For each ES, the algorithm minimizes the total transition cost across the tree by assigning ancestral states that are most consistent with the LMs observed at the leaves. We implemented the following transition costs:

- Unedited → unedited: 0
- LM → same LM: 0
- Unedited → LM: 0.6
- LM → unedited or different LM: 1

These constraints reflect the assumption that the state of an ES should remain the same unless an LM is installed while allowing for some LM-to-LM transitions to account for errors.

Inference of branch lengths

After reconstructing phylogenies and inferring ancestral states, we used ConvexML (v2.0.0) to estimate branch lengths in the tree (*155*). This maximum likelihood algorithm integrates a statistical model of the editing process with inferred ancestral LM states to efficiently estimate branch lengths. ConvexML assumes that editing occurs at a constant rate, meaning that longer branches–representing ancestral cells that take longer to divide–tend to accumulate more LMs than shorter branches. The agreement between our kinetics data and the expected saturating exponential curve suggests that the constant editing rate assumption holds true for PEtracer data for B16-F10 cells in vitro and for 4T1 both in vitro and in vivo (fig. S4E and table S10).

To generate the final phylogenies, we collapsed all branches that were not supported by a LM. This step introduced multifurcations in the tree, reflecting cases where the relationships between descendant cells could not be fully resolved, such as when one or more cell divisions were not marked by an LM. To perform this collapse, we conducted a reverse depth-first traversal of the tree's edges. For each edge, if the inferred LM state of a parent node matched that of its child (i.e., no new LMs were acquired along that branch), we removed the edge and directly connected the child to its grandparent. This approach simplified the tree structure, preserving only branches supported by the available lineage tracing data.

Branch lengths were rescaled to reflect the tracing duration for each experiment, and node depths—defined as the number of edges from the root—were calculated for the collapsed tree. For each phylogeny, the "average tree depth" was determined by computing the mean depth of its leaves.

Plotting phylogenetic trees

Phylogenetic trees were stored using TreeData (v0.1.3, https://github.com/YosefLab/treedata) and visualized with Pycea (v0.0.1, https://github.com/YosefLab/pycea). The plots display inferred branch lengths, with internal nodes sorted by subtree size. When intBCs are shown, they are ordered based on the maximum frequency of LMs at the corresponding ESs, positioning earlier edits on the left. High-confidence clades were manually selected and numbered according to their location in the phylogeny. Note that similar clade colors do not necessarily indicate close phylogenetic relationships due to multifurcations and arbitrary sorting of internal nodes.

Down-sampling analysis of phylogenetic reconstruction accuracy

To assess the impact of detection rate and number of ESs on the accuracy of phylogenetic reconstruction, we performed a down-sampling analysis on the scRNA-seq lineage tracing data with ground-truth barcodes. For each clone, we simulated detection rates ranging from 0.1 to 1 (in increments of 0.1) by randomly dropping out LMs to achieve the target detection rate. Similarly, we simulated different numbers of ESs ranging from 5 to 45 (in increments of 5) by randomly removing ESs. Phylogenies were then reconstructed using neighbor joining, and

the FMI was calculated for puromycin and blasticidin barcode groups to quantify reconstruction accuracy. If the target detection rate or number of ESs exceeded the detection rate or number of ESs for that clone no down-sampling was performed and the true detection rate or number of ESs was reported. Each simulation was repeated 10 times per parameter combination, with FMI scores for each iteration reported in tables S19 and S20 and mean FMI scores reported in the text.

Comparison of pairwise phylogenetic, character, and spatial distance

To compare various distance metrics between cells in a phylogeny, 20,000 random cell pairs were selected, unless the total number of possible pairs was fewer than 20,000, in which case all pairs were used. The distances between each pair were calculated as follows:

1) *Spatial distance:* The Euclidean distance between the x-y coordinates of the two cells. For phylogenies spanning multiple sections, pairs from different sections were excluded.

2) *LM distance:* The Hamming distance between the set of LMs for each cell in the pair, as described in the "Distance-based tree reconstruction" section.

3) *Phylogenetic distance:* The LCA of the cell pair was identified using breadth-first search, and the total path length between the LCA and each of the two cells computed. We use inferred branch lengths for this calculation, so path lengths represent the estimated number of days since the LCA divided.

In vitro 4T1 colony sample preparation and imaging

4T1 cells with high numbers of LTCv2 integrated LTCs and a 24-mer pegArray with no protospacer mismatches that edits to completion within 5 to 7 days were generated as described above ["Generation of 4T1 PEtracer cells edited to completion (fully edited cells)"]. These cells were infected with PEmax-P2A-GFP lentivirus as described above for 1 integrant per cell on day 0 and sorted 48 hours later for high expression of the mCherry-containing tracing cassettes, consistent expression of the BFP-linked pegArray, and for consistent expression of GFP-linked PEmax. Representative sorting gates shown in fig. S16. After sorting, 11,000 actively-editing 4T1 cells were seeded on cleaned, silanized, and PDL-coated 40-mm-diameter round coverslips in 60-mm-diameter petri-dishes with RPMI media as described for fully edited in vitro cell samples above.

After 6 days, cells were fixed by treating with freshly made 4% paraformaldehyde in 1x PBS for 10 min prior to 3x washes with 1x PBS and storage in 1x PBS supplemented with 0.2% RNase Inhibitor Murine (NEB, M0314L) for up to 8 weeks. Samples were prepared for FISH-based lineage readout as described in the "Sample preparation for in-gel T7 amplification" section and then imaged and analyzed as described above. Representative colonies of varying sizes were imaged across 262 FOVs of one coverslip.

Phylogenetic reconstruction of in vitro 4T1 colonies

Segmented nuclei were grouped into spatially distinct colonies using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm ($\varepsilon = 60$ μm) (*156*), with colonies containing fewer than 25 cells filtered out. Since colonies could originate from multiple single cells that landed next to each other and therefore consist of more than one clone, clone calling was performed within each colony as described in the "Clone calling using intBCs" section. The only difference between scRNA-seq and imaging-based clone calling was that spectral clustering was used for initial assignments instead of non-negative matrix factorization (NMF) due to the lower proportion of doublets in imaging data.

LM quality control and phylogenetic tree reconstruction using neighbor joining were performed as described above. However, based on thresholds determined through in silico benchmarking and down-sampling of scRNA-seq lineage tracing data, only nuclei with >60% intBC detection were included in tree reconstruction with neighbor joining. In some cases, phylogenies contained multiple clones that could not be distinguished by the clone-calling algorithm due to similar intBC compositions. These phylogenies were manually split by clone based on visual inspection of LM data: If no single LM was shared by all cells within a phylogeny, major clades marked by distinct LMs were separated into independent phylogenies.

Branch length estimation was performed as described in the "Inference of branch lengths" section and comparison of phylogenetic and spatial distances performed as described in the "Comparison of pairwise phylogenetic, character, and spatial distance" section. Statistics including number of cells and intBC detection rate for the 64 clones are listed in table S20.

Generation of primary 4T1 tumors

4T1 cells engineered with PEmax editor and LTCs (LTCv0) were generated as described above ("epegRNA mismatch library profiling to tune lineage tracing kinetics") and barcoded using low MOI lentiviral infection with an sgRNA library that co-expresses PuroR-T2A-BFP (Addgene #1000000090) as described above ("Lentiviral production" and "Generation of 4T1 PEtracer cells with high numbers of integrated tracing cassettes"). Cells were selected with puromycin (2 μg/ml) for 5 days starting 72 hours after lentiviral infection. Five hundred thousand or 1 million cells were subcutaneously injected into the mammary fat pad of female Balbc/J mice. Following injection, mice were monitored in accordance with IACUC guidelines prior to sacrifice at 6 weeks postinjection. Mice were sacrificed by CO_2 asphyxiation and primary tumor, liver, and lung tissue collected.

scRNA-seq of primary 4T1 tumors

To inform the design of MERFISH gene panels, we generated scRNA-seq libraries from cancer, immune, and stromal cells isolated from primary 4T1 tumors as well as the lungs and livers of two tumor-bearing mice. Tissue was dissociated into single cell suspension by mincing followed by enzymatic digestion with 300 U/ml collagenase/100 U/ml hyaluronidase (STEMCELL Technologies) and 150 μg/ml DNase I (Sigma Aldrich) in RPMI 1640 (Thermo Fisher Scientific) at 37°C with rotation for 25 min (tumor) or 20 min (lung/liver). Homogenized tissue was smashed through a 70-μm sterile strainer and incubated with ACK Lysing Buffer (Thermo Fisher Scientific) for 5 min at room temperature. Single-cell suspensions were stained in flow buffer (2% fetal bovine serum in 1x PBS) with the following antibodies (1:500 dilution):

- anti-CD45 (clone 30-F11, Alexa Fluor 700, Biolegend Cat. No. 103127, RRID AB_493714)
- anti-CD11b (clone M1/70, APC, Biolegend Cat. No. 101211, RRID AB_312794)
- anti-CD11c (clone N418, APC/Cyanine7, Biolegend Cat. No. 117323, RRID AB_830646)
- anti-TCRb (clone H57-597, PE/Cyanine7, Biolegend Cat. No. 109221, RRID AB_893627)
- anti-Ly-6G/Ly-6C (Gr-1) (clone RB6-8C5, Brilliant Violet 711, Biolegend Cat. No. 108443, RRID AB_2562549)
- anti-CD19 (clone 6D5, Brilliant Violet 605, Biolegend Cat. No. 115539, RRID AB_2563067)

Cells were stained for viability using eF506 dye (1:1000, eBioscience). Cells were sorted on BD FACS Aria II for viable cancer cells ($mCherry^+$ $CD45^-$), stromal cells ($mCherry^-$ $CD45^-$), neutrophils ($CD45^+$ $Gr\text{-}1^+$) and other immune cells ($CD45^+$ $Gr\text{-}1^-$). After sorting, cells were pooled together by tissue for three total captures and captured using the 10X Genomics Chromium Next GEM Single Cell 3′ Kit v3.1. scRNA-seq libraries were prepared following the manufacturer's instructions and sequenced on a NovaSeq 6000 (Illumina).

scRNA-seq analysis of primary 4T1 tumors

scRNA-seq data was processed with Cellranger as described above ("scRNA-seq data processing"). Filtered gene-cellBC matrices that

contained only cellBCs with UMI counts that passed the threshold for cell detection were used for further analysis. Additional analysis was performed in R (v.4.2.3) using Seurat (*157*) (v.4.3.0) with default function parameters unless otherwise noted. Doublets were predicted using DoubletFinder (*158*) (v.2.0.3). Cell types were predicted using SingleR (*159*) (v.2.0.0) based on mouse bulk RNA-seq reference data (MouseRNAseqData) from celldex (v.1.8.0). Cells with <200 genes detected, >10% mitochondrial RNA content, or predicted doublets from DoubletFinder and cells annotated as Erythrocytes by SingleR were excluded from analysis. We normalized and identified variable features using the Seurat functions NormalizeData and FindVariableFeatures. We then regressed out percent mitochondrial RNA content and number of UMIs and scaled data with Seurat's ScaleData and ran PCA using variable features with RunPCA. Clusters were identified using shared-nearest-neighbor-based clustering based on the first 10 PCs with resolution = 0.2 with FindNeighbors and FindClusters. The same principal components were used to generate the UMAP projections with RunUMAP. Cell clustering for fine-resolution cell types was performed as described above for malignant/stromal, myeloid, and T cell subsets with the following parameters:

Malignant/stromal: dims = 1:10, resolution = 0.25
Myeloid: dims = 1:15, resolution = 0.25
T cells: dims = 1:10, resolution = 0.3

For cell subsets, cell cycle scoring was performed using Seurat's CellCycleScoring and S phase/G2M phase scores were regressed out during data scaling. Cell types were assigned using SingleR annotations and manually refined based on expression of known marker genes.

In vivo PEtracer sample generation

4T1 cells with high numbers of LTCv2 integrated LTCs and a 24-mer pegArray with protospacer mismatches designed to tune editing to ~70% completion over a 4- to 6-week timescale were generated as described above ["Generation of 4T1 PEtracer cells edited to completion (fully edited cells)"]. These cells were infected with lentivirus expressing PEmax-P2A-GFP on day 0 of the experiment with an infection rate of ~30% and thus roughly one integrant per cell as above and sorted on day 2 for high mCherry signal, consistent BFP signal, and high GFP signal. Representative sorting gates shown in fig. S16. For mouse 1, these cells were expanded in culture for an additional week prior to a second sort on day 9 of the experiment. These cells were again expanded in vitro prior to injection into mice on day 12 of the experiment. For mouse 2 and 3, higher cell numbers were infected on day 0, an initial sort was performed on day 2 and the second sort was performed on day 5 prior to injection into animals on day 6.

Between 200,000 and 500,000 cells undergoing evolving tracing were injected into the tail vein of female Balbc/J x SELECTIV mice (mouse 1) or Balbc/J x [SELECTIV x CMV-Cre] mice (mouse 2 and mouse 3). Following injection, mice were monitored in accordance with IACUC guidelines prior to sacrifice either at a predetermined time point or when mouse health deteriorated to a point that humane sacrifice was necessary (mouse 1 = 39 days postinjection, mouse 2 = 28 days postinjection, mouse 3 = 35 days postinjection). Experimental endpoints were chosen to minimize morbidity due to excessive tumor burden while generating tumors of sufficient size for imaging and to fully utilize the recording capacity of the PEtracer system. Mice were sacrificed by CO_2 asphyxiation prior to perfusion with 1% diethyl pyrocarbonate (DEPC) in sterile PBS. Following perfusion, mouse lungs were dissected out, incubated with 1% DEPC in sterile PBS at 4°C for 3 hours, washed 2x with fresh 1x PBS, and soaked in optimal cutting compound (OCT) for 5 to 10 min at room temperature prior to degassing in a vacuum chamber. Degassed, OCT-embedded samples were frozen on dry ice and stored at −80°C until sectioning.

scRNA-seq analysis of 4T1 in vitro heterogeneity

To characterize in vitro transcriptional heterogeneity and clonal dynamics of the cells used to generate in vivo tumors, we performed scRNA-seq and data processing as described in the "scRNA-seq of fully edited 4T1 cells" and "scRNA-seq data processing" sections. Processed data were analyzed using Scanpy (v1.10.0) (*160*) following standard best practices: Data were normalized and log-transformed, 2000 highly variable genes were identified, and the data were scaled and reduced to 20 principal components (PCs). A 2D embedding was then generated using UMAP (min_dist = 1). Clustering was performed using the Leiden algorithm (resolution = 0.1), which identified two transcriptional clusters. Differential gene expression between clusters was assessed using the rank_genes_groups function (method = *t* test), restricted to genes with mean expression >0.1 across cells. EMT-related genes were highlighted using the EMTome database (http://www.emtome.org/) (*161*).

In vivo MERFISH sample preparation

Frozen lungs were sectioned around −15 to −20°C on a cryostat (Leica CM3050s). We started collecting adjacent 20-μm tissue sections (10 to 20 total sections) once tumors were visible during cryosectioning. For large tumors, we performed multiple collections of tissue sections at different depths (trimming 100 to 200 μm between collections) to sample the entirety of the tumor volume. Tissue slices were collected on silanized and PDL-treated 40-mm coverslips, and fixed by treating with freshly made 4% paraformaldehyde in 1x PBS for 15 min and were washed three times with 1x PBS and stored in 70% ethanol at 4°C for at least 18 hours to permeabilize cell membranes. Tissue sections that did not directly proceed to MERFISH hybridization were stored in 70% ethanol at 4°C for no longer than 2 months. For each collection of sections, we selected a representative section with good morphology and tumor representation for MERFISH imaging and lineage readout to identify tumors with optimal editing kinetics that were selected for additional imaging. Sections with successful MERFISH and lineage imaging were included in further analysis.

Design of MERFISH encoding probes

To discriminate transcriptionally distinct cell types in 4T1 tumor and probe the related cell physiology, we assembled the 124-gene MERFISH panel from two categories:

1) Differentially expressed genes calculated from our annotated 10X scRNA-seq results. A Wilcoxon test was performed to identify the most significant differentially expressed genes.

2) Manually selected genes including canonical cell type markers.

For these candidate genes, we identified all possible 30-nt probe binding sites within annotated transcripts and kept candidate probe sequences that satisfy the following criteria using the MERFISH probe designer (https://github.com/xingjiepan/MERFISH_probe_design): 68°C<Tm<83°C given the salt concentration in 2x SSC buffer; 40%<GC content<70%; no 15-mer match with rRNA and tRNA; transcriptome specificity >99%. After filtering probe binding sites, genes with fewer than 400 possible probe binding sites (roughly 500 nt in transcript length) were removed.

For the selected genes, we designed a codebook by importing a covering design (https://ljcr.dmgordon.org/cover/links.html with parameters: v=16 or 18, k=4, t=3), and then converting this design into a Hamming distance 4, Hamming weight 6 N-bit Hamming code (N=16 for the 124-gene MERFISH library, N=18 for the 175-gene MERFISH library) by removing codewords with less than 4 Hamming distance to other codewords. We then assigned the Hamming codes to genes by balancing the cell type averaged UMI counts from reference scRNA-seq data across all bits and all cell types.

We then followed standard MERFISH library design procedures to generate the probe library. For each gene, we first assembled candidate probes by appending three readout sequences to the probe targeting sequences (reverse-complement of the probe targeting sequences designed above). These three readouts were randomly chosen from four assigned readouts assigned to this gene as indicated in the codebook

and an extra "A" was added in between readout sequences and probe targeting sequences. Next, we appended 20-nt forward and reverse primers to the 5′ and 3′ of these candidate probes. We then selected 90 to 96 of the overlapping probes for each gene among all possible candidate probes while minimizing the overlap binding to the target transcripts between neighboring probes. Then these fully assembled probes were further screened for cross-hybridization and any probe sharing >17 nt of reverse-complement sequence with other probes was dropped. Full probe library sequences could be found in table S21.

To improve cell type resolution and monitor tumor progression we assembled the 175-gene MERFISH library from the following categories:

1) Genes from our initial 124-gene MERFISH library after removing genes with no obvious spatial or cell type specific enrichment (110 genes)
2) Cell cycle marker genes for identifying cell cycle states (7 genes)
3) EMT markers (5 genes)
4) Hypoxia markers (2 genes)
5) Additional FGF pathway markers (3 genes)
6) Finer cell type markers from existing references (34 genes)
7) Hits from a CRISPR screen in 4T1 cells (114) (4 genes)

Probes for this final set of selected genes were filtered with the same design criteria as the 124-gene library described above. For this 175-gene MERFISH library we used 18 choose 4 Hamming codes. Full probe library sequences could be found in table S25.

MERFISH encoding probe hybridization and tissue clearing

Both in vitro and in vivo samples were washed once with 2x SSC prior to equilibration in encoding-probe wash buffer (30% formamide in 2x SSC supplemented with 0.1% Tween-20) for 10 to 15 minutes at room temperature. Encoding-probe wash buffer was then aspirated from the coverslip, and the coverslip was inverted onto a 50-ul droplet of probe mixture on a parafilm-covered 60-mm diameter petri dish. The probe mixture comprised 1 nM of each encoding probe for MERFISH imaging, approximately 10 nM of each encoding probe for the two-color mCherry single-molecule FISH imaging, and 1 μM of a polyA-anchor probe (/5Acryd/TTGAGTGGATGGAGTGTAATT+TT+TT+TT+TT+TT+TT+TT+TT+TT+T, where T+ represents a locked nucleic acid T nucleobase, and /5Acryd/ is a 5′ acrydite modification available through IDT) in 2x SSC with 30% v/v formamide, 0.1% w/v yeast tRNA (approximately, Life Technologies) and 10% v/v dextran sulfate (D8906, Sigma). We then incubated the sample at 37°C for 36 to 48 hours. The polyA-anchor probe was hybridized to polyA sequences on polyadenylated endogenous mRNAs, thereby anchoring these RNAs to the polyacrylamide gel. After hybridization, the samples were washed in the encoding-probe wash buffer for 30 min at 47°C for a total of two times to remove excess encoding probes and polyA-anchor probes. The samples were washed once with 2x SSC prior to poly-acrylamide gel-embedding.

Samples were embedded and tissue cleared similarly to section "Sample preparation for in-gel T7 amplification" with fully-edited 4T1 cells" with the following modifications. Following tissue clearing, samples were washed in 2x SSC buffer supplemented with 0.2% (v/v) Rnase Inhibitor Murine (NEB) at room temperature for 2 hours. We changed the buffer every 20 min to ensure sufficient washing. Samples could be stored at 2x SSC supplemented with 0.2% Rnase Inhibitor Murine at 4°C for up to 2 weeks.

MERFISH Imaging

MERFISH imaging was performed for each sample as described above in the section "Imaging of integration barcodes and lineage marks" with the following changes: Two adaptors and two readout probes were used in each round of imaging; z-stack images consisting of 24 or 25 planes separated by 1.2 μm were collected for all channels; the number of imaging rounds was adjusted for each MERFISH library, specifically, for the 124-gene MERFISH library we used 16 bits (eight imaging rounds with 2 bits per round) and for the 175-gene MERFISH library we used 18 bits (nine imaging rounds with 2 bits per round). Four sections were imaged for mice 1 and 3, and 1 section was imaged for mouse 2. The number of fields of view and the set of tumors captured for each section is listed in table S22.

MERFISH decoding

All MERFISH image analysis was performed using MERlin (available at https://github.com/zhengpuas47/MERlin) (*162*), as previously described (*152*). First, we identified the locations of fiducial beads in each FOV in each round of imaging and used these locations to determine the x–y drift in the stage position relative to the first round of imaging and to align images for each FOV across all imaging rounds. We then high-pass filtered the MERFISH image stacks for each FOV to remove background, performed 10 rounds of Lucy-Richardson deconvolution to tighten RNA spots, and low-pass filtered them to account for small movements in the centroid of RNAs between imaging rounds. Individual RNA molecules imaged by MERFISH were identified by our previously published pixel-based decoding algorithm using MERlin (*162*). After assigning barcodes to each pixel independently, we aggregated adjacent pixels that were assigned with the same barcodes into putative RNA molecules, and then filtered the list of putative RNA molecules to enrich for correctly identified transcripts as previously described for a gross barcode misidentification rate at 5% using MERlin.

Nuclei segmentation for in vivo samples

Nuclei segmentation for in vivo samples was performed using Cellpose, following a similar approach to that used for in vitro nuclei segmentation (see "Nuclei segmentation for in vitro samples" section for reference) (*98*). However, modifications were necessary to accurately segment the small, densely packed nuclei observed in vivo. To enhance image sharpness and contrast of DAPI images before segmentation, we applied the Deconwolf deconvolution algorithm (*99*). Deconwolf was run for 100 iterations, with each FOV divided into four tiles to facilitate computation on an Nvidia GeForce GTX 2080ti GPU. The resulting deconvolved images were segmented using Cellpose (v3.1.0) with nucleus diameter set to 17 pixels. Nuclei with a volume <100 μm^3 were excluded from analysis. The number of nuclei detected per tumor section is listed in table S22.

Assignment of MERFISH-decoded transcripts to cells

Using the 3D coordinates from MERlin, transcripts were assigned to nuclei following the same procedure as for T7 amplicons (see "Assignment of T7 amplicons to cells"). However, rather than assigning cytoplasmic transcripts to the nearest nucleus, we used Proseg (v1.1.3) to infer the most likely cell boundaries based on the assumption that cytoplasmic gene expression mirrors that of the nucleus (*102*). Proseg was run separately for each tissue section with the following parameters: voxel-layers = 3, ncomponents = 10, nbglayers = 10, enforce-connectivity = True, max-transcript-nucleus-distance = 10, and nuclear-reassignment-prob = 0.2. The cells identified by Proseg were then mapped back to nuclear masks based on their assigned transcripts. Each Proseg cell was assigned to the nucleus with the highest Jaccard similarity, provided a similarity >0.4, which helped filter out spurious Proseg cell calls. Transcripts outside of nuclei were assigned using this mapping if the Proseg assignment probability was >0.5, while transcripts already within nuclei retained their original assignments. The total number of assigned transcripts and mean number of transcripts per cell for each tumor section is listed in table S22.

MERFISH cell clustering and annotation

RNA count matrices were used for cell clustering and annotation with resolVI (*100*), a scvi-tools (*163*) (version 1.3.0) method for deconvolution and label transfer of spatial transcriptomics data. Separate models were trained for the 124-gene and 175-gene MERFISH libraries using sections M1 S3 and M3 S2, respectively. For model training, initial cell clustering was performed with Scanpy (v1.10.0) following standard

preprocessing steps: Cells with fewer than five genes or 20 transcripts were removed, data were normalized and log-transformed, total transcript counts were regressed out, and the data were scaled and reduced in dimensionality (20 PCs). Clusters were then identified using the Leiden algorithm (resolution = 0.6), and UMAP embeddings (min_dist = 0.25) were generated for visualization. Fine resolution clustering was achieved by re-clustering individual clusters with varying resolutions (0.1 to 0.4).

Labeled RNA count matrices were used as input for the resolVI model, which was trained for 100 epochs under a semi-supervised framework without down-sampling counts. The trained model was used for reference mapping with the full dataset from all sections imaged using the same MERFISH library, with input data filtered, normalized, regressed, and scaled as described, and then projected. Cells were annotated by predicting labels and retaining those with a probability >0.5. Diffusion and background transcript proportions were calculated, and cells with >20% diffusion or background transcripts were removed. The latent representation from resolVI served as input for a final UMAP projection (min_dist = 0.2).

For cells profiled with the 175-gene MERFISH library cell cycle phase was classified using the "score_genes_cell_cycle" function implemented by Scanpy (v1.10.0). *Dscc1*, *Msh2*, *Rad51*, *Rpa2* and *Ung* were used to score S phase, while *Cdca2*, *Kif2c*, *Ncapd2*, *Nek2* were used to score G2/M.

Comparison of MERFISH and scRNA-seq data

To assess MERFISH data quality, we compared correlation between MERFISH replicates as well as the correlation between MERFISH, bulk RNA-seq and scRNA-seq. For comparison between MERFISH replicates, we computed the average counts per cell for each transcript and calculated the Pearson correlation between two sections (M1 S1 and M1 S2). For comparison between MERFISH and bulk RNA-seq, we obtained normalized counts per million (CPM) data from published 4T1-derived lung metastasis bulk RNA-seq data (*106*) and calculated the Pearson correlation with average counts per cell obtained by MERFISH. To compare to scRNA-seq data described above ("Single-cell RNA analysis of primary 4T1 tumors"), we computed the average normalized counts per cell per scRNA-seq or MERFISH cluster and calculated the pairwise Pearson correlation between cell types, including all cell types that were identified in both datasets.

Cell type neighborhood enrichment and density analysis

Cell neighborhood enrichment analysis was performed using Squidpy (v1.6.2) and Hotspot (v1.1.1) (*107*). First, spatial neighborhoods were computed using Squidpy by identifying the six nearest neighbors of each cell based on 2D Euclidean distance, then neighborhood enrichment was assessed with the "nhood_enrichment" function. We then used Hotspot to identify cell type neighborhoods using sample M1 S2. Spatially variable genes were selected based on Geary's *C* autocorrelation (calculated with 100 neighbors), retaining genes with Geary's *C* >0.04 and an average expression >0.1 counts per cell. Hotspot was then run with a minimum_gene_threshold=5 and fdr_threshold=0.01. For each section, scores for the five neighborhoods identified in M1 S2 were calculated using the same parameters. Each cell was assigned to the neighborhood with the highest score.

Tumor and normal lung boundaries were delineated through a series of geometric operations and represented as polygons using geopandas (v1.0.1). To construct tumor boundaries, malignant cell centroids were dilated, and the outer edge of their combined area was determined. In contrast, normal lung boundaries were defined by subtracting the tumor boundary polygons from the union of dilated AT1/AT2 and Club cell centroids.

For each cell, distances to both tumor and lung boundaries were determined using the geopandas "sjoin_nearest" function. The distance to the lung-tumor interface was defined differently depending on cell location: For cells within the tumor boundary, it was the negative of the lung boundary distance; for cells outside the tumor boundary, it was the distance to the tumor boundary. These distances were scaled per tumor in each section by dividing by the maximum interface distance observed.

Cell type densities (cells/mm^2) were determined by counting the total number of cells of each type within a 0.1-mm radius circle centered on each cell and dividing by the area of that circle. Rolling averages of these densities, computed along the gradient from normal lung to the outer tumor edge, were calculated using windows of 2000 cells then scaled by dividing by the maximum average density for each cell type.

FISH-based lineage readout after MERFISH for in vivo samples

After MERFISH imaging, samples were removed from FCS2 flow chamber, washed once with hybridization wash buffer (2x SSC and 30% v/v formamide), further digested, washed, and T7 in situ amplified as described in the "Sample preparation for in-gel T7 amplification" section. Samples were then imaged and analyzed as described above.

Sample alignment between MERFISH and lineage imaging

To align samples on the home-built imaging platform for both MERFISH and lineage readouts, during MERFISH imaging we captured DAPI images using the 10x objective across entire tissue sections. During the setup of lineage imaging, we first adjusted the orientation of samples manually to roughly align their orientations with that used for MERFISH imaging. We then captured a new DAPI mosaic image for each complete tissue section using the 10x objective. Between the DAPI images captured for each sample during MERFISH and lineage imaging, we manually picked 10 positions as featured anchoring points and recorded their x-y coordinates. From these paired coordinates we performed a singular value decomposition (SVD) of the multiplicated position matrix to acquire a rotation and translation matrix by assuming samples only experienced affine transformation. Given the rotation and translation that we calculated, we applied the transformation to the center coordinates of the FOVs imaged during MERFISH for each sample and used these transformed coordinates as the new FOV coordinates for lineage imaging.

After image acquisition, we corrected residual alignment issues and sample distortions computationally. We generated large mosaic images by combining DAPI images from all fields of view acquired during the MERFISH and lineage readouts. These mosaics were down-sampled by a factor of 4, smoothed with an unsharp mask (Gaussian sigma = five pixels), intensity-rescaled by clipping values >98th percentile, and rotated by an angle determined using anchoring points. Residual x-y drift was corrected by using phase cross-correlation to calculate the shift between the mosaics, and the lineage mosaic was adjusted accordingly. To address non-linear distortions, TV-L1 optical flow (implemented in scikit-image v0.24.0) (*164*) was used to compute a vector field aligning the two mosaics. This vector field was divided into 100 by 100 pixel tiles, and for each tile the mean, intensity-weighted shift was calculated. The global x-y shift was then added to the local shift for each tile, and the tiles were saved. In experiments combining MERFISH and lineage imaging, nuclear masks were generated from the MERFISH images, and T7 amplicons were assigned to these masks after adjusting their 3D coordinates based on the corresponding tile.

Generation, preparation, and imaging of primary tumors with fully edited 4T1 cells

Primary tumors were generated as above with fully edited 4T1 cells used for in vitro validation experiments. One million cells were subcutaneously injected into the mammary fat pad of female Balbc/J mice. Following injection, mice were monitored in accordance with IACUC guidelines prior to sacrifice at 2.5 weeks postinjection. Mice were sacrificed by CO_2 asphyxiation and tumors collected.

Primary tumors embedded in OCT were prepared as in section "In vivo MERFISH sample preparation"; specifically, the 20-μm sections were placed on treated coverslips and treated as MERFISH samples but skipping MERFISH imaging procedures. Samples were prepared for FISH-based lineage readout as described in the "Sample preparation

for in-gel T7 amplification" section and then imaged and analyzed as described above. In total, 146 FOVs were imaged, capturing the entirety of one primary tumor in one section.

Processing of lineage data from in vivo tumor sections

T7 amplicons in lineage readout images were detected, quantified, and decoded as described above, with one key modification in the intBC decoding step. Instead of using a whitelist of all 2171 possible intBCs, we limited it to the 36 intBCs found in the tumor-initiating clones. This smaller whitelist improved decoding under higher in vivo noise levels and allowed us to increase the maximum allowable distance between the codeword and the intensity profile to 2, thanks to the increased Hamming distances among the 36 intBCs.

LM decoding and quality control were carried out as described earlier, using a logistic regression classifier trained on fully edited in vitro cells. For the primary tumor containing fully edited cells, clones were identified based on their Jaccard similarity to intBCs and LMs in the fully edited clone whitelist (see "Quantification of lineage tracing data accuracy in fully edited cells" section for details). Cells with a Jaccard similarity <0.2 were not assigned to a clone, as they likely represent non-malignant cells lacking LTCs. LM decoding accuracy was then evaluated for malignant cells, serving as an out-of-sample validation of the classifier trained on in vitro data.

To account for imperfect nuclei segmentation and sample distortion between the MERFISH and lineage imaging, decoded intBCs were assigned to a whitelist of malignant cells derived from the MERFISH data. If the position of the decoded T7 amplicon was just outside of a malignant cell's nuclear segmentation mask, it was assigned to that cell even if it was closer to a non-malignant cell, as long as the distance to the nuclear mask was less than 4 μm.

Reconstruction of in vivo tumor phylogenies

Spatial data from all tissue sections for each mouse were manually aligned in the x-y plane using tissue morphology. Cells were then grouped into spatially distinct tumors using the DBSCAN clustering algorithm ($\varepsilon = 0.3$ mm), with abutting tumors separated manually. Clone calling, as described in the "Clone calling using intBCs" section, was performed for each tumor to determine its clonality and the corresponding set of intBCs. As with the colony data, spectral clustering was used for clone calling instead of NMF, due to the lower proportion of doublets.

Phylogenies for malignant cells were reconstructed using neighbor joining, as detailed in the "Distance-based tree reconstruction" section, incorporating data from all sections for each tumor, except for one mouse 3 section which was excluded because of low intBC detection. For tumor 2 in mouse 2 (M2 T2), separate phylogenies were built for the two contributing clones. To account for partial nuclei generated by tissue sectioning, only nuclei with centroids located more than 5 μm from the top or bottom of the section were used. Among these intact nuclei, the intBC detection rate was calculated, and only cells with a detection rate >60% were used for tree reconstruction. To improve tree reconstruction with missing data, ESs with LMs shared by all cells were excluded, and select ESs with LMs defining early splits were manually upweighted by a factor of 2 before calculating the Hamming distance matrix. Statistics for the 12 reconstructed tumor phylogenies are reported in table S23.

Further analysis was performed on phylogenies with edit fractions >50% (M1 T1, M1 T2, M1 T4, M2 T5, and M1 T1). This included inference of ancestral LM states and branch lengths as described in the "Inference of branch lengths" section and comparison of phylogenetic, character, and spatial distance as described in the "Comparison of pairwise phylogenetic, character, and spatial distance" section. To compute the LM installation efficiency and account for edits that occur at different depths of the tree, we computed the overall transition probability from the unedited state to a given LM based on transitions along each branch in the tree. The resulting LM installation efficiency is normalized by tumor and ES to account for differences in LM installation rate.

Linking in vitro 4T1 cells to in vivo tumors

To investigate how in vitro heterogeneity in PEtracer 4T1 cells relates to their potential to form tumors in vivo, we mapped in vitro cells (see "scRNA-seq analysis of 4T1 in vitro heterogeneity" section) to individual tumors based on their detected intBCs. For each in vitro cell, the 30-nt sequencing barcode was matched to its corresponding 183-nt imaging barcode, as described in the "scRNA-seq of fully edited 4T1 cells" section. We then calculated the Jaccard similarity between these barcodes and the barcode whitelist associated with each tumor.

Reconstruction of tumor growth dynamics

To reconstruct the tumor growth in 3D format we estimated the z-position of sections for mouse 1 tumor 1 (M1 T1) based on section number and thickness, and then inferred the position of each phylogenetic clade as the mean 3D position of cells within that clade. The inferred tumor origin was determined by averaging the 3D positions of all cells in the phylogeny, and the positions of ancestral nodes between the origin and each clade's common ancestor were inferred by averaging the positions of their descendants with the inferred tumor origin.

To estimate the number of extant cells in each clade during the growth of M1 T1, we divided the timeline from day 0 to 39 into 20 discrete time points. At each time point, we cut the phylogenetic tree at the corresponding depth and counted the number of branches for each clade. These branches represent ancestral cells that were alive at that time point.

Calculation of local LM diversity

We define local character diversity as the average pairwise LM distance between all cells within a 100-μm radius of a given cell. To compute this metric, we first identify spatial neighbors based on x-y coordinates and then extract the corresponding $M \times M$ subset from the $N \times N$ weighted Hamming distance matrix given the M nearest neighbors. The mean pairwise distance within this subset represents the local character diversity. Cells in phylogenetically diverse regions have higher local character diversity scores, as neighboring cells are less likely to share similar LMs.

Estimation of cell fitness

Phylogenetic fitness was estimated using the infer_fitness function from the Jungle package (https://github.com/felixhorns/jungle) which implements the probabilistic fitness estimation algorithm described by Neher *et al.* 2014 (*165*). Conceptually, this algorithm leverages two key assumptions:

1) *Fitter cells divide more rapidly*, producing more offspring and leading to shorter branch lengths.

2) *Fitness changes gradually along lineages*, meaning a cell's fitness is expected to be similar to that of its ancestors and relatives.

Phylogenetic fitness was calculated using the inferred branch lengths generated by ConvexML, as described in the "Branch length inference" section.

As an alternative tree-independent fitness metric, we calculated the mean neighbor LM distance for each cell in the phylogeny. Using the *N-by-N* weighted Hamming distance matrix, computed as described in the "Distance-based tree reconstruction" section, we identified the 20 nearest neighbors of each cell and calculated the mean LM distance among them. This metric is based on the expectation that fitter cells will have closely related neighbors with similar LMs, as there has been less time for additional LMs to accumulate compared to less fit cells.

Fitness correlation analysis

To identify gene expression patterns, spatial positions, and cell type neighborhoods associated with increased phylogenetic fitness, we computed the Pearson correlation between fitness and each feature for

every tumor with fitness estimates. For gene expression, raw counts were used; for spatial positioning, we employed distance to the tumor boundary and the lung-tumor interface; and for cell type neighborhoods, we used local cell type density (see "Cell type neighborhood enrichment and density analysis" for details on boundary distance and density calculations). Additionally, we calculated the Pearson correlation between these features and the distance to the tumor boundary to separate the effects of cellular features from those related to tumor position. *P* values were derived using a two-sided Student's *t* test and adjusted for multiple comparisons with the Benjamini-Hochberg procedure. Results from this analysis are reported in table S24.

Comparison of MERFISH libraries

To compare gene expression data generated by our 124- and 175-gene MERFISH libraries, we computed the average normalized counts per cell per annotated cluster across all samples profiled using each library and calculated the pairwise Pearson correlation between cell types, including all cell types that were identified using both libraries.

Cancer cell module analysis

Hotspot (v1.1.1) was used to call spatially coherent gene expression modules within the mouse 3 malignant cell population (*107*). Modules were identified using sample M3 S2. Spatially variable genes were selected based on Geary's C autocorrelation (calculated with 100 neighbors), retaining genes with Geary's C >0.05 and an average expression >0.1 counts per cell. Hotspot was then run with a minimum_gene_threshold=4 and fdr_threshold=0.01. For each M3 section, scores for the four modules identified in S2 were calculated using the same parameters. Each cell was assigned to the module with the highest score.

To identify cell types associated with each module, we calculated the mean density (see "Cell type neighborhood enrichment and density analysis" for reference) of each cell type across cells in each module, and then z-scored these values. For genes selected by Hotspot to distinguish between modules, we calculated their mean expression by aggregating normalized counts across cells in each module, and then z-scored these values.

Transition probabilities between Hotspot modules were calculated by first identifying the set of phylogenetic neighbors for each cell—those with a last common ancestor within 10 days. Then, for each module, we computed the fraction of these neighbors assigned to every other module, yielding the transition probabilities between modules.

Moran's *I* statistic for heritability and spatial variation

We use Moran's *I* autocorrelation statistic to quantify the heritability and spatial variation of gene expression values and cancer cell module scores within a phylogeny. While Moran's *I* is commonly applied to spatial data to measure the clustering or dispersion of a variable, it can also be used in phylogenetics by replacing the spatial neighbor graph with a phylogenetic neighbor graph (*166*). Moran's *I* is defined as:

$$I = \frac{N^2}{\sum_{i,j=1}^{N} w_{i,i}} \times \frac{\sum_{i,j=1}^{N} w_{i,i} x_i x_j}{\sum_{i=1}^{N} x_i^2}$$

where x is the z-scored variable of interest, $w_{i,j}$ represents the connectivity between cells i and j ($w_{i,jj} = 1$ for neighbors and 0 otherwise), and N is the number of cells. To construct the phylogenetic neighbor graph, each cell in the phylogeny is connected to all other cells that share a LCA within the last 10 days. Similarly, to construct a spatial neighbor graph, each malignant cell is connected to all other malignant cells within 100 μm. Moran's *I* is then computed for z-scored MERFISH counts and cancer cell module scores. *P* values are calculated with a permutation test (1000 permutations) and adjusted for multiple comparisons with the Benjamini-Hochberg procedure (*167*). Results from this analysis are reported in table S26.

REFERENCES AND NOTES

1. T. Stadler, O. G. Pybus, M. P. H. Stumpf, Phylodynamics for cell biologists. *Science* **371**, eaah6266 (2021). doi: 10.1126/science.aah6266; pmid: 33446527
2. S. Domcke, J. Shendure, A reference cell tree will serve science better than a reference cell atlas. *Cell* **186**, 1103–1114 (2023). doi: 10.1016/j.cell.2023.02.016; pmid: 36931241
3. D. E. Wagner, A. M. Klein, Lineage tracing meets single-cell omics: Opportunities and challenges. *Nat. Rev. Genet.* **21**, 410–427 (2020). doi: 10.1038/s41576-020-0223-2; pmid: 32235876
4. J. E. Sulston, E. Schierenberg, J. G. White, J. N. Thomson, The embryonic cell lineage of the nematode Caenorhabditis elegans. *Dev. Biol.* **100**, 64–119 (1983). doi: 10.1016/0012-1606(83)90201-4; pmid: 6684600
5. C. Qiu *et al.*, A single-cell time-lapse of mouse prenatal development from gastrula to birth. *Nature* **626**, 1084–1093 (2024). doi: 10.1038/s41586-024-07069-w; pmid: 38355799
6. C. Qiu *et al.*, Systematic reconstruction of cellular trajectories across mouse embryogenesis. *Nat. Genet.* **54**, 328–341 (2022). doi: 10.1038/s41588-022-01018-x; pmid: 35288709
7. L. Wolpert, Positional information and the spatial pattern of cellular differentiation. *J. Theor. Biol.* **25**, 1–47 (1969). doi: 10.1016/S0022-5193(69)80016-0; pmid: 4390734
8. A. Chen *et al.*, Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792.e21 (2022). doi: 10.1016/j.cell.2022.04.003; pmid: 35512705
9. R. J. Metzger, O. D. Klein, G. R. Martin, M. A. Krasnow, The branching programme of mouse lung development. *Nature* **453**, 745–750 (2008). doi: 10.1038/nature07005; pmid: 18463632
10. V. G. Sankaran, J. S. Weissman, L. I. Zon, Cellular barcoding to decipher clonal dynamics in disease. *Science* **378**, eabm5874 (2022). doi: 10.1126/science.abm5874; pmid: 36227997
11. C. A. Lareau *et al.*, Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. *Nat. Biotechnol.* **39**, 451–461 (2021). doi: 10.1038/s41587-020-0645-6; pmid: 32788668
12. B. Treutlein *et al.*, Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371–375 (2014). doi: 10.1038/nature13173; pmid: 24739965
13. M. G. Jones, D. Yang, J. S. Weissman, New Tools for Lineage Tracing in Cancer In Vivo. *Annu. Rev. Cancer Biol.* **7**, 111–129 (2023). doi: 10.1146/annurev-cancerbio-061421-123301
14. N. McGranahan, C. Swanton, Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* **168**, 613–628 (2017). doi: 10.1016/j.cell.2017.01.018; pmid: 28187284
15. N. Zahir, R. Sun, D. Gallahan, R. A. Gatenby, C. Curtis, Characterizing the ecological and evolutionary dynamics of cancer. *Nat. Genet.* **52**, 759–767 (2020). doi: 10.1038/s41588-020-0668-4; pmid: 32719518
16. M. Dhainaut *et al.*, Spatial CRISPR genomics identifies regulators of the tumor microenvironment. *Cell* **185**, 1223–1239.e20 (2022). doi: 10.1016/j.cell.2022.02.015; pmid: 35290801
17. T. Baslan *et al.*, Ordered and deterministic cancer genome evolution after p53 loss. *Nature* **608**, 795–802 (2022). doi: 10.1038/s41586-022-05082-5; pmid: 35978189
18. S. Y. Ku *et al.*, Rb1 and Trp53 cooperate to suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. *Science* **355**, 78–83 (2017). doi: 10.1126/science.aah4199; pmid: 28059767
19. P. Mu *et al.*, SOX2 promotes lineage plasticity and antiandrogen resistance in TP53- and RB1-deficient prostate cancer. *Science* **355**, 84–88 (2017). doi: 10.1126/science.aah4307; pmid: 28059768
20. J. M. Chan *et al.*, Lineage plasticity in prostate cancer depends on JAK/STAT inflammatory signaling. *Science* **377**, 1180–1191 (2022). doi: 10.1126/science.abn0478; pmid: 35981096
21. S. VanHorn, S. A. Morris, Next-Generation Lineage Tracing and Fate Mapping to Interrogate Development. *Dev. Cell* **56**, 7–21 (2021). doi: 10.1016/j.devcel.2020.10.021; pmid: 33217333
22. M. Al Bakir *et al.*, The evolution of non-small cell lung cancer metastases in TRACERx. *Nature* **616**, 534–542 (2023). doi: 10.1038/s41586-023-05729-x; pmid: 37046095
23. A. M. Frankell *et al.*, The evolution of lung cancer and impact of subclonal selection in TRACERx. *Nature* **616**, 525–533 (2023). doi: 10.1038/s41586-023-05783-5; pmid: 37046096
24. L. S. Ludwig *et al.*, Lineage Tracing in Humans Enabled by Mitochondrial Mutations and Single-Cell Genomics. *Cell* **176**, 1325–1339.e22 (2019). doi: 10.1016/j.cell.2019.01.022; pmid: 30827679
25. A. Sottoriva *et al.*, A Big Bang model of human colorectal tumor growth. *Nat. Genet.* **47**, 209–216 (2015). doi: 10.1038/ng.3214; pmid: 25665006
26. I. Martincorena *et al.*, High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880–886 (2015). doi: 10.1126/science.aaa6806; pmid: 25999502
27. E. Mitchell *et al.*, Clonal dynamics of haematopoiesis across the human lifespan. *Nature* **606**, 343–350 (2022). doi: 10.1038/s41586-022-04786-y; pmid: 35650442

28. D. Biswas *et al.*, A clonal expression biomarker associates with lung cancer mortality. *Nat. Med.* **25**, 1540–1548 (2019). doi: 10.1038/s41591-019-0595-z; pmid: 31591602
29. R. Rosenthal *et al.*, Neoantigen-directed immune escape in lung cancer evolution. *Nature* **567**, 479–485 (2019). doi: 10.1038/s41586-019-1032-7; pmid: 30894752
30. T. H. H. Coorens *et al.*, Extensive phylogenies of human development inferred from somatic mutations. *Nature* **597**, 387–392 (2021). doi: 10.1038/s41586-021-03790-y; pmid: 34433963
31. S. Park *et al.*, Clonal dynamics in early human embryogenesis inferred from somatic mutation. *Nature* **597**, 393–397 (2021). doi: 10.1038/s41586-021-03786-8; pmid: 34433967
32. M. Spencer Chapman *et al.*, Lineage tracing of human development through somatic mutations. *Nature* **595**, 85–90 (2021). doi: 10.1038/s41586-021-03548-6; pmid: 33981037
33. E. Tzouanacou, A. Wegener, F. J. Wymeersch, V. Wilson, J.-F. Nicolas, Redefining the progression of lineage segregations during mammalian embryogenesis by clonal analysis. *Dev. Cell* **17**, 365–376 (2009). doi: 10.1016/j.devcel.2009.08.002; pmid: 19758561
34. K. K. Chow *et al.*, Imaging cell lineage with a synthetic digital recording system. *Science* **372**, eabb3099 (2021). doi: 10.1126/science.abb3099; pmid: 33833095
35. J. Livet *et al.*, Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**, 56–62 (2007). doi: 10.1038/nature06293; pmid: 17972876
36. H. J. Snippert *et al.*, Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* **143**, 134–144 (2010). doi: 10.1016/j.cell.2010.09.016; pmid: 20887898
37. A. Rodrigo Albors, P. A. Halley, K. G. Storey, Lineage tracing of axial progenitors using Nkx1-2CreER^{T2} mice defines their trunk and tail contributions. *Development* **145**, dev164319 (2018). doi: 10.1242/dev.164319; pmid: 30201686
38. C. Weinreb, A. Rodriguez-Fraticelli, F. D. Camargo, A. M. Klein, Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* **367**, eaaw3381 (2020). doi: 10.1126/science.aaw3381; pmid: 31974159
39. M. M. Chan *et al.*, Molecular recording of mammalian embryogenesis. *Nature* **570**, 77–82 (2019). doi: 10.1038/s41586-019-1184-5; pmid: 31086336
40. K. L. Frieda *et al.*, Synthetic recording and in situ readout of lineage information in single cells. *Nature* **541**, 107–111 (2017). doi: 10.1038/nature20777; pmid: 27869821
41. A. McKenna *et al.*, Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016). doi: 10.1126/science.aaf7907; pmid: 27229144
42. D. M. Chadly, K. L. Frieda, C. Gui, L. Klock, M. Tran, M. Y. Sui, Y. Takei, R. Bouckaert, C. Lois, L. Cai, M. B. Elowitz, Reconstructing cell histories in space with image-readable base editor recording. bioRxiv 2024.01.03.573434 [Preprint] (2024). https://doi.org/10.1101/2024.01.03.573434.
43. T. B. Loveless *et al.*, Lineage tracing and analog recording in mammalian cells by single-site DNA writing. *Nat. Chem. Biol.* **17**, 739–747 (2021). doi: 10.1038/s41589-021-00769-8; pmid: 33753928
44. T. B. Loveless *et al.*, Open-ended molecular recording of sequential cellular events into DNA. *Nat. Chem. Biol.* **21**, 512–521 (2025). doi: 10.1038/s41589-024-01764-5; pmid: 39543397
45. S. Bowling *et al.*, An Engineered CRISPR-Cas9 Mouse Line for Simultaneous Readout of Lineage Histories and Gene Expression Profiles in Single Cells. *Cell* **181**, 1410–1422.e27 (2020). doi: 10.1016/j.cell.2020.04.048; pmid: 32413320
46. L. Li *et al.*, A mouse model with high clonal barcode diversity for joint lineage, transcriptomic, and epigenomic profiling in single cells. *Cell* **186**, 5183–5199.e22 (2023). doi: 10.1016/j.cell.2023.09.019; pmid: 37852258
47. J. Choi *et al.*, A time-resolved, multi-symbol molecular recorder via sequential genome editing. *Nature* **608**, 98–107 (2022). doi: 10.1038/s41586-022-04922-8; pmid: 35794474
48. B. Raj *et al.*, Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* **36**, 442–450 (2018). doi: 10.1038/nbt.4103; pmid: 29608178
49. B. Spanjaard *et al.*, Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat. Biotechnol.* **36**, 469–473 (2018). doi: 10.1038/nbt.4124; pmid: 29644996
50. A. Alemany, M. Florescu, C. S. Baron, J. Peterson-Maduro, A. van Oudenaarden, Whole-organism clone tracing using single-cell sequencing. *Nature* **556**, 108–112 (2018). doi: 10.1038/nature25969; pmid: 29590089
51. R. Kalhor *et al.*, Developmental barcoding of whole mouse via homing CRISPR. *Science* **361**, eaat9804 (2018). doi: 10.1126/science.aat9804; pmid: 30093604
52. Z. He *et al.*, Lineage recording in human cerebral organoids. *Nat. Methods* **19**, 90–99 (2022). doi: 10.1038/s41592-021-01344-8; pmid: 34969984
53. J. J. Quinn *et al.*, Single-cell lineages reveal the rates, routes, and drivers of metastasis in cancer xenografts. *Science* **371**, eabc1944 (2021). doi: 10.1126/science.abc1944; pmid: 33479121
54. D. Yang *et al.*, Lineage tracing reveals the phylodynamics, plasticity, and paths of tumor evolution. *Cell* **185**, 1905–1923.e25 (2022). doi: 10.1016/j.cell.2022.04.015; pmid: 35523183
55. M. G. Jones *et al.*, Spatiotemporal lineage tracing reveals the dynamic spatial architecture of tumor growth and metastasis. bioRxiv 2024.10.21.619529 (2024). https://doi.org/10.1101/2024.10.21.619529
56. A. Askary *et al.*, In situ readout of DNA barcodes and single base edits facilitated by in vitro transcription. *Nat. Biotechnol.* **38**, 66–75 (2020). doi: 10.1038/s41587-019-0299-4; pmid: 31740838
57. D. Bressan, G. Battistoni, G. J. Hannon, The dawn of spatial omics. *Science* **381**, eabq4964 (2023). doi: 10.1126/science.abq4964; pmid: 37535749
58. K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, X. Zhuang, Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015). doi: 10.1126/science.aaa6090; pmid: 25858977
59. Y. Goltsev *et al.*, Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* **174**, 968–981.e15 (2018). doi: 10.1016/j.cell.2018.07.010; pmid: 30078711
60. J.-R. Lin *et al.*, Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *eLife* **7**, e31657 (2018). doi: 10.7554/eLife.31657; pmid: 29993362
61. G. Gut, M. D. Herrmann, L. Pelkmans, Multiplexed protein maps link subcellular organization to cellular states. *Science* **361**, eaar7042 (2018). doi: 10.1126/science.aar7042; pmid: 30072512
62. E. Lubeck, A. F. Coskun, T. Zhiyentayev, M. Ahmad, L. Cai, Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods* **11**, 360–361 (2014). doi: 10.1038/nmeth.2892; pmid: 24681720
63. A. Rao, D. Barkley, G. S. França, I. Yanai, Exploring tissue architecture using spatial transcriptomics. *Nature* **596**, 211–220 (2021). doi: 10.1038/s41586-021-03634-9; pmid: 34381231
64. X. Zhuang, Spatially resolved single-cell genomics and transcriptomics by imaging. *Nat. Methods* **18**, 18–22 (2021). doi: 10.1038/s41592-020-01037-8; pmid: 33408406
65. X. Wang *et al.*, Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018). doi: 10.1126/science.aat5691; pmid: 29930089
66. A. V. Anzalone *et al.*, Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019). doi: 10.1038/s41586-019-1711-4; pmid: 31634902
67. P. J. Chen *et al.*, Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* **184**, 5635–5652.e29 (2021). doi: 10.1016/j.cell.2021.09.018; pmid: 34653350
68. P. J. Chen, D. R. Liu, Prime editing for precise and highly versatile genome manipulation. *Nat. Rev. Genet.* **24**, 161–177 (2023). doi: 10.1038/s41576-022-00541-1; pmid: 36344749
69. A. V. Anzalone, L. W. Koblan, D. R. Liu, Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **38**, 824–844 (2020). doi: 10.1038/s41587-020-0561-9; pmid: 32572269
70. M. van Overbeek *et al.*, DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Mol. Cell* **63**, 633–646 (2016). doi: 10.1016/j.molcel.2016.06.037; pmid: 27499295
71. M. W. Shen *et al.*, Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* **563**, 646–651 (2018). doi: 10.1038/s41586-018-0686-x; pmid: 30405244
72. M. G. Jones *et al.*, Inference of single-cell phylogenies from lineage tracing data using Cassiopeia. *Genome Biol.* **21**, 92 (2020). doi: 10.1186/s13059-020-02000-8; pmid: 32290857
73. D. F. Robinson, L. R. Foulds, Comparison of phylogenetic trees. *Math. Biosci.* **53**, 131–147 (1981). doi: 10.1016/0025-5564(81)90043-2
74. D. E. Critchlow, D. K. Pearl, C. Qian, The Triples Distance for Rooted Bifurcating Phylogenetic Trees. *Syst. Biol.* **45**, 323–334 (1996). doi: 10.1093/sysbio/45.3.323
75. R. Wang, R. Zhang, A. Khodaverdian, N. Yosef, Theoretical guarantees for phylogeny inference from single-cell lineage tracing. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2203352120 (2023). doi: 10.1073/pnas.2203352120; pmid: 36927151
76. N. Saitou, M. Nei, The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987). pmid: 3447015
77. R. R. Sokal, C. D. Michener, *A Statistical Method for Evaluating Systematic Relationships* (University of Kansas, 1958).
78. P. T. L. C. Clausen, Scaling neighbor joining to one million taxa with dynamic and heuristic neighbor joining. *Bioinformatics* **39**, btac774 (2023). doi: 10.1093/bioinformatics/btac774; pmid: 36453849
79. R. J. Ihry *et al.*, p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat. Med.* **24**, 939–946 (2018). doi: 10.1038/s41591-018-0050-6; pmid: 29892062
80. E. Haapaniemi, S. Botla, J. Persson, B. Schmierer, J. Taipale, CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* **24**, 927–930 (2018). doi: 10.1038/s41591-018-0049-z; pmid: 29892067
81. M. L. Leibowitz *et al.*, Chromothripsis as an on-target consequence of CRISPR-Cas9 genome editing. *Nat. Genet.* **53**, 895–905 (2021). doi: 10.1038/s41588-021-00838-7; pmid: 33846636
82. M. Kosicki, K. Tomberg, A. Bradley, Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018). doi: 10.1038/nbt.4192; pmid: 30010673

83. O. M. Enache *et al.*, Cas9 activates the p53 pathway and selects for p53-inactivating mutations. *Nat. Genet.* **52**, 662–668 (2020). doi: 10.1038/s41588-020-0623-4; pmid: 32424350
84. G. Yu *et al.*, Prediction of efficiencies for diverse prime editing systems in multiple cell types. *Cell* **186**, 2256–2272.e23 (2023). doi: 10.1016/j.cell.2023.03.034; pmid: 37119812
85. J. W. Nelson *et al.*, Engineered pegRNAs improve prime editing efficiency. *Nat. Biotechnol.* **40**, 402–410 (2022). doi: 10.1038/s41587-021-01039-7; pmid: 34608327
86. S. Bae, J. Park, J.-S. Kim, Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014). doi: 10.1093/bioinformatics/btu048; pmid: 24463181
87. S. Levesque *et al.*, Marker-free co-selection for successive rounds of prime editing in human cells. *Nat. Commun.* **13**, 5909 (2022). doi: 10.1038/s41467-022-33669-z; pmid: 36207338
88. P. Datlinger *et al.*, Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297–301 (2017). doi: 10.1038/nmeth.4177; pmid: 28099430
89. C. J. Aslakson, F. R. Miller, Selective events in the metastatic process defined by analysis of the sequential dissemination of subpopulations of a mouse mammary tumor. *Cancer Res.* **52**, 1399–1405 (1992). pmid: 1540948
90. D. L. Dexter *et al.*, Heterogeneity of tumor cells from a single mouse mammary tumor. *Cancer Res.* **38**, 3174–3181 (1978). pmid: 210930
91. I. J. Fidler, Biological behavior of malignant melanoma cells correlated to their survival in vivo. *Cancer Res.* **35**, 218–224 (1975). pmid: 1109790
92. T. Lu, C. E. Ang, X. Zhuang, Spatially resolved epigenomic profiling of single cells in complex tissues. *Cell* **185**, 4448–4464.e17 (2022). doi: 10.1016/j.cell.2022.09.035; pmid: 36272405
93. J. R. Moffitt *et al.*, High-performance multiplexed fluorescence in situ hybridization in culture and tissue with matrix imprinting and clearing. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14456–14461 (2016). doi: 10.1073/pnas.1617699113; pmid: 27911841
94. E. B. Fowlkes, C. L. Mallows, A Method for Comparing Two Hierarchical Clusterings. *J. Am. Stat. Assoc.* **78**, 553–569 (1983). doi: 10.1080/01621459.1983.10478008
95. A. Alard *et al.*, Breast cancer cell mesenchymal transition and metastasis directed by DAP5/eIF3d-mediated selective mRNA translation. *Cell Rep.* **42**, 112646 (2023). doi: 10.1016/j.celrep.2023.112646; pmid: 37314929
96. M. C. P. Smith *et al.*, CXCR4 regulates growth of both primary and metastatic breast cancer. *Cancer Res.* **64**, 8604–8612 (2004). doi: 10.1158/0008-5472.CAN-04-1844; pmid: 15574767
97. B. K. Martin *et al.*, Optimized single-nucleus transcriptional profiling by combinatorial indexing. *Nat. Protoc.* **18**, 188–207 (2023). doi: 10.1038/s41596-022-00752-0; pmid: 36261634
98. C. Stringer, M. Pachitariu, Cellpose3: One-click image restoration for improved cellular segmentation. *Nat. Methods* **22**, 592–599 (2025). doi: 10.1038/s41592-025-02595-5; pmid: 39939718
99. E. Wernersson *et al.*, Deconwolf enables high-performance deconvolution of widefield fluorescence microscopy images. *Nat. Methods* **21**, 1245–1256 (2024). doi: 10.1038/s41592-024-02294-7; pmid: 38844629
100. C. Ergen, N. Yosef, ResolVI: Addressing noise and bias in spatial transcriptomics. bioRxiv 2025.01.20.634005 [Preprint] (2025). https://doi.org/10.1101/2025.01.20.634005.
101. P. Cadinu *et al.*, Charting the cellular biogeography in colitis reveals fibroblast trajectories and coordinated spatial remodeling. *Cell* **187**, 2010–2028.e30 (2024). doi: 10.1016/j.cell.2024.03.013; pmid: 38569542
102. D. C. Jones *et al.*, Cell simulation as cell segmentation. *Nat. Methods* **22**, 1331–1342 (2025). doi: 10.1038/s41592-025-02697-0; pmid: 40404994
103. R. Zilionis *et al.*, Single-Cell Transcriptomics of Human and Mouse Lung Cancers Reveals Conserved Myeloid Populations across Individuals and Species. *Immunity* **50**, 1317–1334.e10 (2019). doi: 10.1016/j.immuni.2019.03.009; pmid: 30979687
104. M. Guo *et al.*, Guided construction of single cell reference for human and mouse lung. *Nat. Commun.* **14**, 4566 (2023). doi: 10.1038/s41467-023-40173-5; pmid: 37516747
105. B. Schrörs *et al.*, Multi-Omics Characterization of the 4T1 Murine Mammary Gland Tumor Model. *Front. Oncol.* **10**, 1195 (2020). doi: 10.3389/fonc.2020.01195; pmid: 32793490
106. C. M. Ferrer *et al.*, The glutathione S-transferase Gstt1 drives survival and dissemination in metastases. *Nat. Cell Biol.* **26**, 975–990 (2024). doi: 10.1038/s41556-024-01426-7; pmid: 38862786
107. D. DeTomaso, N. Yosef, Hotspot identifies informative gene modules across modalities of single-cell genomics. *Cell Syst.* **12**, 446–456.e9 (2021). doi: 10.1016/j.cels.2021.04.005; pmid: 33951459
108. M. Casanova-Acebes *et al.*, Tissue-resident macrophages provide a pro-tumorigenic niche to early NSCLC cells. *Nature* **595**, 578–584 (2021). doi: 10.1038/s41586-021-03651-8; pmid: 34135508
109. D. Laoui *et al.*, Tumor hypoxia does not drive differentiation of tumor-associated macrophages but rather fine-tunes the M2-like macrophage population. *Cancer Res.* **74**, 24–30 (2014). doi: 10.1158/0008-5472.CAN-13-1196; pmid: 24220244
110. A.-T. Henze, M. Mazzone, The impact of hypoxia on tumor-associated macrophages. *J. Clin. Invest.* **126**, 3672–3679 (2016). doi: 10.1172/JCI84427; pmid: 27482883
111. C. Murdoch, M. Muthana, C. E. Lewis, Hypoxia regulates macrophage functions in inflammation. *J. Immunol.* **175**, 6257–6263 (2005). doi: 10.4049/jimmunol.175.10.6257; pmid: 16272275
112. C. Lewis, C. Murdoch, Macrophage responses to hypoxia: Implications for tumor progression and anti-cancer therapies. *Am. J. Pathol.* **167**, 627–635 (2005). doi: 10.1016/S0002-9440(10)62038-X; pmid: 16127144
113. A. Dongre, R. A. Weinberg, New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nat. Rev. Mol. Cell Biol.* **20**, 69–84 (2019). doi: 10.1038/s41580-018-0080-4; pmid: 30459476
114. X. Wang *et al.*, In vivo CRISPR screens identify the E3 ligase Cop1 as a modulator of macrophage infiltration and cancer immunotherapy target. *Cell* **184**, 5357–5374.e22 (2021). doi: 10.1016/j.cell.2021.09.006; pmid: 34582788
115. L. Moses, L. Pachter, Museum of spatial transcriptomics. *Nat. Methods* **19**, 534–546 (2022). doi: 10.1038/s41592-022-01409-2; pmid: 35273392
116. S. G. Rodriques *et al.*, Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019). doi: 10.1126/science.aaw1219; pmid: 30923225
117. A. J. C. Russell *et al.*, Slide-tags enables single-nucleus barcoding for multimodal spatial genomics. *Nature* **625**, 101–109 (2024). doi: 10.1038/s41586-023-06837-4; pmid: 38093010
118. R. R. Stickels *et al.*, Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat. Biotechnol.* **39**, 313–319 (2021). doi: 10.1038/s41587-020-0739-1; pmid: 33288904
119. S. Y. Tam, V. W. C. Wu, H. K. W. Law, Hypoxia-Induced Epithelial-Mesenchymal Transition in Cancers: HIF-1α and Beyond. *Front. Oncol.* **10**, 486 (2020). doi: 10.3389/fonc.2020.00486; pmid: 32322559
120. N. Ferrara, VEGF and the quest for tumour angiogenesis factors. *Nat. Rev. Cancer* **2**, 795–803 (2002). doi: 10.1038/nrc909; pmid: 12360282
121. Y. Zhang, R. A. Weinberg, Epithelial-to-mesenchymal transition in cancer: Complexity and opportunities. *Front. Med.* **12**, 361–373 (2018). doi: 10.1007/s11684-018-0656-6; pmid: 30043221
122. T. Biancalani *et al.*, Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat. Methods* **18**, 1352–1362 (2021). doi: 10.1038/s41592-021-01264-7; pmid: 34711971
123. T. Abdelaal, S. Mourragui, A. Mahfouz, M. J. T. Reinders, SpaGE: Spatial Gene Enhancement using scRNA-seq. *Nucleic Acids Res.* **48**, e107 (2020). doi: 10.1093/nar/gkaa740; pmid: 32955565
124. W. E. Allen, T. R. Blosser, Z. A. Sullivan, C. Dulac, X. Zhuang, Molecular and spatial signatures of mouse brain aging at single-cell resolution. *Cell* **186**, 194–208.e18 (2023). doi: 10.1016/j.cell.2022.12.010; pmid: 36580914
125. R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, A. Regev, Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015). doi: 10.1038/nbt.3192; pmid: 25867923
126. R. Fang *et al.*, Three-dimensional single-cell transcriptome imaging of thick tissues. *eLife* **12**, RP90029 (2024). doi: 10.7554/eLife.90029; pmid: 39727221
127. G. Chu, U. Mai, H. Schmidt, B. J. Raphael, Maximum likelihood inference of time-scaled cell lineage trees with mixed-type missing data using LAML. *Genome Biol.* **26**, 189 (2025). doi: 10.1186/s13059-025-03649-9; pmid: 40604857
128. S.-W. Wang, M. J. Herriges, K. Hurley, D. N. Kotton, A. M. Klein, CoSpar identifies early cell fate biases from single-cell transcriptomic and lineage information. *Nat. Biotechnol.* **40**, 1066–1074 (2022). doi: 10.1038/s41587-022-01209-1; pmid: 35190690
129. N. Konno *et al.*, Deep distributed computing to reconstruct extremely large lineage trees. *Nat. Biotechnol.* **40**, 566–575 (2022). doi: 10.1038/s41587-021-01111-2; pmid: 34992246
130. W. Fang *et al.*, Quantitative fate mapping: A general framework for analyzing progenitor state dynamics via retrospective lineage barcoding. *Cell* **185**, 4604–4620.e32 (2022). doi: 10.1016/j.cell.2022.10.028; pmid: 36423582
131. R. A. Saunders *et al.*, Perturb-Multimodal: A platform for pooled genetic screens with imaging and sequencing in intact mammalian tissue. *Cell* **0**, S0092-8674(25)00572-0 (2025). doi: 10.1016/j.cell.2025.05.022; pmid: 40513557
132. W. Chen *et al.*, Symbolic recording of signalling and cis-regulatory element activity to DNA. *Nature* **632**, 1073–1081 (2024). doi: 10.1038/s41586-024-07706-4; pmid: 39020177
133. A. Tanay, A. Regev, Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**, 331–338 (2017). doi: 10.1038/nature21350; pmid: 28102262
134. Tabula Muris Consortium, Single-cell transcriptomics of 20 mouse organs creates a *Tabula Muris*. *Nature* **562**, 367–372 (2018). doi: 10.1038/s41586-018-0590-4; pmid: 30283141
135. S. R. Srivatsan *et al.*, Embryo-scale, single-cell spatial transcriptomics. *Science* **373**, 111–117 (2021). doi: 10.1126/science.abb9536; pmid: 34210887
136. A. Sampath Kumar *et al.*, Spatiotemporal transcriptomic maps of whole mouse embryos at the onset of organogenesis. *Nat. Genet.* **55**, 1176–1185 (2023). doi: 10.1038/s41588-023-01435-6; pmid: 37414952
137. X. Huang *et al.*, Single-cell, whole-embryo phenotyping of mammalian developmental disorders. *Nature* **623**, 772–781 (2023). doi: 10.1038/s41586-023-06548-w; pmid: 37968388

138. R. C. Jones *et al.*, The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science* **376**, eabl4896 (2022). doi: 10.1126/science.abl4896; pmid: 35549404
139. D. A. Cusanovich *et al.*, A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* **174**, 1309–1324.e18 (2018). doi: 10.1016/j.cell.2018.06.052; pmid: 30078704
140. O. Rozenblatt-Rosen, M. J. T. Stubbington, A. Regev, S. A. Teichmann, The Human Cell Atlas: From vision to reality. *Nature* **550**, 451–453 (2017). doi: 10.1038/550451a; pmid: 29072289
141. A. Regev *et al.*, Science Forum: The Human Cell Atlas. *eLife* **6**, e27041 (2017). doi: 10.7554/eLife.27041; pmid: 29206104
142. K. J. Travaglini *et al.*, A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619–625 (2020). doi: 10.1038/s41586-020-2922-4; pmid: 33208946
143. J. Zengel *et al.*, Hardwiring tissue-specific AAV transduction in mice through engineered receptor expression. *Nat. Methods* **20**, 1070–1081 (2023). doi: 10.1038/s41592-023-01896-x; pmid: 37291262
144. K. Clement *et al.*, CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019). doi: 10.1038/s41587-019-0032-3; pmid: 30809026
145. J. M. Replogle *et al.*, Maximizing CRISPRi efficacy and accessibility with dual-sgRNA libraries and optimal effectors. *eLife* **11**, e81856 (2022). doi: 10.7554/eLife.81856; pmid: 36576240
146. J. N. Zadeh *et al.*, NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.* **32**, 170–173 (2011). doi: 10.1002/jcc.21596; pmid: 20645303
147. C. Xia, J. Fan, G. Emanuel, J. Hao, X. Zhuang, Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 19490–19499 (2019). doi: 10.1073/pnas.1912459116; pmid: 31501331
148. Q. Xu, M. R. Schlabach, G. J. Hannon, S. J. Elledge, Design of 240,000 orthogonal 25mer DNA barcode probes. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 2289–2294 (2009). doi: 10.1073/pnas.0812506106; pmid: 19171886
149. R. Schmidt *et al.*, CRISPR activation and interference screens decode stimulation responses in primary human T cells. *Science* **375**, eabj4008 (2022). doi: 10.1126/science.abj4008; pmid: 35113687
150. A. J. Hill *et al.*, On the design of CRISPR-based single-cell molecular screens. *Nat. Methods* **15**, 271–274 (2018). doi: 10.1038/nmeth.4604; pmid: 29457792
151. J.-H. Su, P. Zheng, S. S. Kinrot, B. Bintu, X. Zhuang, Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell* **182**, 1641–1659.e26 (2020). doi: 10.1016/j.cell.2020.07.032; pmid: 32822575
152. M. Zhang *et al.*, Molecularly defined and spatially resolved cell atlas of the whole mouse brain. *Nature* **624**, 343–354 (2023). doi: 10.1038/s41586-023-06808-9; pmid: 38092912
153. J. R. Moffitt *et al.*, High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 11046–11051 (2016). doi: 10.1073/pnas.1612826113; pmid: 27625426
154. D. Sankoff, Minimal Mutation Trees of Sequences. *SIAM J. Appl. Math.* **28**, 35–42 (1975). doi: 10.1137/0128004
155. S. Prillo, A. Ravoor, N. Yosef, Y. S. Song, Convex ML: Scalable and accurate inference of single-cell chronograms from CRISPR/Cas9 lineage tracing data. bioRxiv 2023.12.03.569785 [Preprint] (2023). https://doi.org/10.1101/2023.12.03.569785.
156. M. Ester, H.-P. Kriegel, J. Sander, X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise" in *KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, Portland, OR, 2 to 4 August 1996 (AAAI Press, 1996), pp. 226–231.
157. Y. Hao *et al.*, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021). doi: 10.1016/j.cell.2021.04.048; pmid: 34062119
158. C. S. McGinnis, L. M. Murrow, Z. J. Gartner, DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst.* **8**, 329–337.e4 (2019). doi: 10.1016/j.cels.2019.03.003; pmid: 30954475
159. D. Aran *et al.*, Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019). doi: 10.1038/s41590-018-0276-y; pmid: 30643263
160. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018). doi: 10.1186/s13059-017-1382-0; pmid: 29409532
161. S. V. Vasaikar *et al.*, EMTome: A resource for pan-cancer analysis of epithelial-mesenchymal transition genes and signatures. *Br. J. Cancer* **124**, 259–269 (2021). doi: 10.1038/s41416-020-01178-9; pmid: 33299129
162. G. Emanuel, seichhorn, H. Babcock, leonardosepulveda, timblosser, ZhuangLab/MERlin: MERlin v0.1.6, Zenodo (2020); https://doi.org/10.5281/zenodo.3758540.
163. A. Gayoso *et al.*, A Python library for probabilistic analysis of single-cell omics data. *Nat. Biotechnol.* **40**, 163–166 (2022). doi: 10.1038/s41587-021-01206-w; pmid: 35132262
164. J. Sánchez Pérez, E. Meinhardt-Llopis, G. Facciolo, TV-L1 Optical Flow Estimation. *Image Process. Line* **3**, 137–150 (2013). doi: 10.5201/ipol.2013.26
165. R. A. Neher, C. A. Russell, B. I. Shraiman, Predicting evolution from the shape of genealogical trees. *eLife* **3**, e03568 (2014). doi: 10.7554/eLife.03568; pmid: 25385532
166. R. Chaligne *et al.*, Epigenetic encoding, heritability and plasticity of glioma transcriptional cell states. *Nat. Genet.* **53**, 1469–1479 (2021). doi: 10.1038/s41588-021-00927-7; pmid: 34594037
167. S. J. Rey, L. Anselin, "PySAL: A Python Library of Spatial Analytical Methods" in *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, M. M. Fischer, A. Getis, Eds. (Springer, 2010), pp. 175–193.doi: 10.1007/978-3-642-03647-7_11
168. W. Colgan, L. Koblan, K. Yost, P. Zheng, PEtracer Data, *Figshare* (2025); https://doi.org/10.6084/m9.figshare.28473866.
169. W. Colgan, K. Yost, P. Zheng, jweissmanlab/PEtracer-2025: v0.0.3, Zenodo (2025); https://doi.org/10.5281/zenodo.15882011.

ACKNOWLEDGMENTS

We thank members of the Weissman, Yosef, and Zhuang labs for helpful discussions. We thank J. Moffitt for helpful discussions and advice on MERFISH sample preparation. We also thank the Whitehead Flow Cytometry Core, the Whitehead Genome Technology Core, the Koch Institute Hope Babette Tang Histology Facility, and the Broad Institute Klarman Cell Observatory. **Funding:** This work was supported by the Helen Hay Whitney/Howard Hughes Medical Institute (HHMI) fellowship and the Eunice Kennedy Shriver NICHD Pathway to Independence Award NIH K99HD118574 (L.W.K.), NIH grant K00CA253729 (K.E.Y.), the Fayez Sarofim Fellowship of the Damon Runyon Cancer Research Foundation (P. Z.), the Damon Runyon Dale Frey Award (D.Y.), the NCI Transition Career Development Award 1K22CA289207 (D.Y.), the NIH Director's New Innovator Award 1DP2OD037078 (D.Y.), the Lung Cancer Research Foundation Leading Edge Award (D.Y.), the NCI Pathway to Independence Award NIH K99CA286968 (M.G.J.), NIH grant 3T32DK007191-50S1 (J.S.), the Jane Coffin Childs Memorial Fund for Medical Research (A.S.), the Harvard Society of Fellows (R.A.S. and W.E.A.), the Burroughs-Wellcome Fund Career Award at the Scientific Interface (W.E.A.), HHMI (X.Z. and J.S.W.), NIH Centers of Excellence in Genomic Science (RM1HG009490) (J.S.W.), Chan Zuckerberg Initiative 2024-346405 (5022) (J.S.W.), the Ludwig Center at MIT (J.S.W.), the Melanoma Research Alliance (C-13384133), (J.S.W.), the Bay Area Cancer Target Discover and Development (14198SC) (J.S.W.), and the Mathers Charitable Foundation (J.S.W.) **Author contributions:** Conceptualization: L.W.K., K.E.Y., J.S.W.; Formal analysis: W.N.C., P.Z., K.E.Y., L.W.K.; Funding acquisition: J.S.W.; Investigation: L.W.K., K.E.Y., P.Z., A.K., J.S., A.S., D.S.; Methodology: L.W.K., K.E.Y., P.Z., W.N.C., W.E.A., D.Y., R.A.S., X.Z.; Software: W.N.C., P.Z., M.G.J., C.E.; Supervision: J.S.W., X.Z., N.Y.; Visualization: W.N.C., L.W.K., K.E.Y., P.Z.; Writing – original draft: L.W.K., K.E.Y., W.N.C., P.Z., J.S.W.; Writing – review & editing: L.W.K., K.E.Y., W.N.C., P.Z., J.S.W., M.G.J., D.Y., A.K., J.S., A.S., D.S., C.E., R.A.S., X.Z., W.E.A., N.Y. **Competing interests:** M.G.J. consults for and has equity in Vevo Therapeutics. K.E.Y. is a consultant for Cartography Biosciences and Curie Bio. D.Y. declares outside interest in DEM Biopharma. W.N.C. consults for Merck Pharmaceuticals. X.Z. is a co-founder and consultant of Vizgen. N.Y. is a consultant for Cytoreason. R.A.S. and W.E.A. are consultants for Xaira Therapeutics. J.S.W. declares outside interest in 5 AM Venture, Amgen, nChroma Bio, DEM Biosciences, KSQ Therapeutics, Maze Therapeutics, Tenaya Therapeutics, Tessera Therapeutics, Thermo Fisher Scientific, Third Rock Ventures, and Xaira Therapeutics. All other authors declare that they have no competing interests. **Data and materials availability:** Plasmids generated in this study are available on Addgene. All scRNA-seq data generated in this study are available under Gene Expression Omnibus (GEO) accession no. GSE290975. All other sequencing data are available on Sequence Read Archive (SRA) under BioProject accession no. PRJNA1231108. Additional processed data including MERFISH data and lineage reconstructions are available on Figshare (*168*). Custom code used in this study is available at Zenodo (*169*). **License information:** Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. This article is subject to HHMI's Open Access to Publications policy. HHMI lab heads have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the author-accepted manuscript (AAM) of this article can be made freely available under a CC BY 4.0 license immediately upon publication.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adx3800
Figs. S1 to S16; Tables S1 to S30; Movie S1; MDAR Reproducibility Checklist
Submitted 12 March 2025; accepted 14 July 2025; published online 24 July 2025
10.1126/science.adx3800

Science
JOURNALS AAAS



Publish your research in the *Science* family of journals

The *Science* family of journals (*Science, Science Advances, Science Immunology, Science Robotics, Science Signaling*, and *Science Translational Medicine)* are among the most highly-regarded journals in the world for quality and selectivity. Our peer-reviewed journals are committed to publishing cutting-edge research, incisive scientific commentary, and insights on what's important to the scientific world at the highest standards.

Submit your research today!

Learn more at **Science.org/journals**

RESEARCH ARTICLE SUMMARY

PIGMENTATION

A mosaic of modular variation at a single gene underpins convergent plumage coloration

Dave Lutgen *et al.*

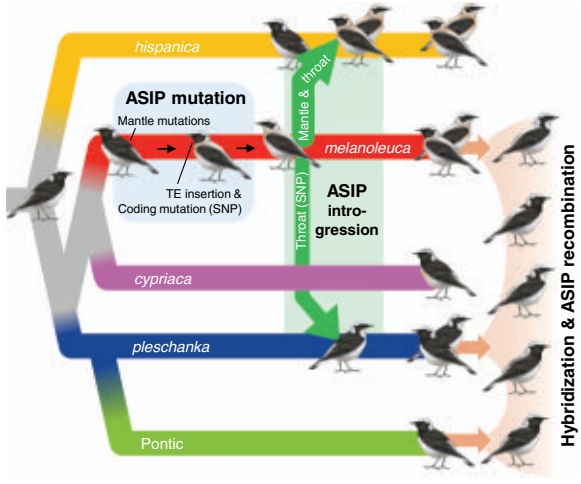
Full article and list of author affiliations:
https://doi.org/10.1126/science.ado8005

INTRODUCTION: Phenotypic diversification abundantly reuses existing variation that is available within species or introgressed from other species. Still, insights into how such variation is integrated with new mutations to form new phenotypes and how this process may be limited by genetic architectures remain scarce.

RATIONALE: In this work, we investigated the genetic architecture and evolutionary history of plumage coloration in a group of songbirds. Wheatears (genus *Oenanthe*) show a high incidence of convergent plumage color evolution and frequent hybridization in the *Oenanthe hispanica* complex, within which three species (*O. hispanica*, *O. melanoleuca*, and *O. pleschanka*) exhibit a black or white throat color polymorphism, and the mantle coloration that is divergent black or white between species is abundantly recombined by hybridization. We leveraged the throat color polymorphism and the recombined mantle coloration to infer the genetic architecture underpinning plumage coloration. On the basis of population genetic patterns observed for the color-related genomic variation, we reconstructed plumage colorations' evolutionary histories.

RESULTS: Genome-wide association studies associate throat and mantle coloration with variation at a single gene, the agouti signaling protein (ASIP) locus. Throat coloration is underpinned by the interaction of a new transposable element (TE) insertion 5′ of ASIP with nonsynonymous variation within the ASIP coding sequence. Mantle coloration is governed by additive effects at multiple sites up to 50 kbp 5′ of ASIP. Expression patterns and isoform structures of ASIP in regrowing feathers confirm this role of ASIP in the regulation of the melanin-based plumage coloration. Population genetic analyses based on the ancestral recombination graph place the evolution of white throat and mantle coloration in *O. melanoleuca*, from where both introgressed into *O. hispanica*, and white throat coloration introgressed into *O. pleschanka*. In *O. melanoleuca* and *O. hispanica*, throat phenotypes are associated with different nitrogen stable isotope signatures that suggest an association of throat coloration with alternative foraging niches. At the genus level, the convergent evolution of throat and mantle coloration involved variation different from that recovered in the *O. hispanica* complex; that is, new variation.

CONCLUSION: The convergent plumage coloration in wheatears evolved from a combination of new and introgressed variation at a single gene. Whereas the convergent pattern of coloration predominantly roots in past introgression within the *O. hispanica* complex, it involved new mutations at genus scale. Introgression further contributed to the formation of recombined phenotypes in extant hybrid zones that may constitute a source for future phenotypic evolution. The complex patterns of color evolution in wheatears demonstrate how phenotypes can be decoupled between body parts within the constraints of a monogenic architecture. Despite the strong linkage within a single gene, fine-scale recombination can integrate linked regulatory modules from divergent origins into new, recombined haplotypes. This appears to imperatively involve long-term recombination; be it within species, in their ancestors, or in hybrid zones that ever increasingly come into focus as notable melting pots that enable evolutionary tinkering. □



Plumage color evolution in wheatears. White mantle and throat coloration evolved in *O. melanoleuca* through regulatory mutations (mantle) and coding mutations and TE insertion (throat) at the ASIP gene. Introgression of ASIP variation led to white mantle coloration in *O. hispanica* and a throat polymorphism in *O. hispanica* and *O. pleschanka*. Fine-scale ASIP recombination in hybrid zones underpins the expression of intermediate plumage phenotypes. SNP, single-nucleotide polymorphism.

Corresponding author: Reto Burri (reto.burri@vogelwarte.ch) Cite this article as D. Lutgen *et al.*, *Science* **390**, eado8005 (2025). DOI: 10.1126/science.ado8005

PIGMENTATION

A mosaic of modular variation at a single gene underpins convergent plumage coloration

Dave Lutgen[1,2,3], Valentina Peona[2,4], Madeline A. Chase[2], Niloofar Alaei Kakhki[5], Fritjof Lammers[2], Stacey G. de Souza[6], Anne-Lyse Ducrest[7], Marta Burri[2], Pavlos Andriopoulos[8], Sifiso M. Lukhele[6], Michaella Moysi[6], Elizabeth Yohannes[2], Abdin Abbasov[9], Tamer Albayrak[10], Mansour Aliabadian[11], Nicolas Auchli[2], Vasileios Bontzorlos[12], Ioulios Christoforou[13], José Luis Copete[14], Egidio Fulco[15], Jesus T. Garcia[16], Zura Javakhishvili[17], Anna Kazazou[18], Fumin Lei[19,20], Yang Liu[21], Nika Paposhvili[17], Robert Patchett[22], Áron Péter[23], Raphael Ritter[24], Attila D. Sándor[23,25,26], Fabian Schneider[27], Petar Shurulinkov[28]†, Sergey Sklyarenko[29], Borut Stumberger[30], Abulfaz Tagiyev[31,32], Alessia Uboldi[2], Nikitas Vogiatzis[18], Fanny Taborsak-Lines[33]‡, Joel Gruselius[34], Liqun Yao[33], Catherine L. Peichel[1], Alexander Suh[35,36,37], Pierre-Alexandre Gagnaire[38], Alexander N. G. Kirschel[6], Manuel Schweizer[39,40], Holger Schielzeth[3,41], Reto Burri[1,2,3]*

The reshuffling of genomic variation from multiple origins is an important contributor to phenotypic diversification, yet insights into the evolutionary trajectories of this combinatorial process and their interplay with genetic architecture remain scarce. We show that convergent plumage color evolution in wheatears involves a monogenic architecture with modular variation introgressed at the agouti signaling protein (ASIP) locus. Introgression of a new transposable element insertion and linked protein-coding variation underpin a transspecific throat color polymorphism, which stable isotopes suggest is associated with alternative foraging niches. Cointrogression of linked regulatory ASIP variation resulted in mantle color convergence in one species, whereas convergent color evolution at the genus level required new variation. Our results demonstrate evolutionary trajectories from introgressed variation realized within the constraints of a monogenic architecture.

Phenotypic evolution was long perceived as a slow process driven by a gradual accumulation of new mutations (*1*). However, recent studies across diverse organisms show that the reshuffling of already existing genetic variation, whether standing within species or introgressed from other species ("preexisting variation"), provides an important fast-track to phenotypic evolution (*2*–*6*). Yet, we still have much to learn about the trajectories along which this reshuffling occurs and how it is shaped by the genetic architecture of phenotypes (*6*).

Prominent examples such as Denisovan-derived adaptations in humans, adaptive introgression in butterflies, Darwin's finches and Yellowstone wolves (*3*, *7*–*9*), and adaptation from ancestral variation in stickleback (*10*) all demonstrate preexisting variation as a potent source for phenotypic evolution. These examples highlight how rapid phenotypic evolution can be facilitated by preexisting variation at single genes. By contrast, the evolution of more complex phenotypes, such as recombined phenotypes composed of reshuffled combinations of multiple traits, typically involve genetic variation spanning multiple genes. For example, the adaptive radiation of cichlid fish and convergent color evolution in several bird genera show how individual phenotypic traits can be recombined into rich suites of phenotypes (*2*, *11*–*15*).

Whether the evolution of phenotypic richness typically involves preexisting variation remains a key question. The answer may in part depend on the genetic architecture of phenotypes. Although single genes repeatedly hit by new mutations over time may generate complex phenotypic patterns within single traits (*16*), examples of complex recombined phenotypes underpinned by monogenic architectures remain scarce (*9*, *17*, *18*). This raises the question of how genetic architecture limits the role of preexisting variation in shaping complex phenotypic diversity. Although the cited examples show the power of combinatorial processes in driving phenotypic and species diversity (*6*), substantial gaps remain in our understanding of the evolutionary trajectories along which the reshuffling of preexisting variation into complex phenotypes occurs. The role of genetic architectures in constraining this combinatorial process warrants further investigation (*19*).

To obtain insights into the role of preexisting variation in the evolution of recombined phenotypes and its interplay with genetic architecture, we investigated the genomic basis and evolutionary trajectories of plumage coloration in a group of songbirds, the wheatears (genus *Oenanthe*). The high incidence of convergent evolution [according to (*20*), including all instances of independent evolution of a phenotype; elsewhere referred to as parallel or repeated evolution (*21*)] and hybridization in wheatears offers powerful opportunities to investigate the involvement of combinatorial processes in the evolution of phenotypic diversity (*22*). Males of most wheatear species display plumages combining black and white feather tracts. The coloration of single feather tracts largely switches independently across species, resulting in phenotypes that reappear multiple times across the genus (*22*–*24*). The convergent evolution of these phenotypes may have involved a combination of ancestral, introgressed, and new genetic variation (*22*). However, to fully understand how these processes contribute to phenotypic evolution, it is essential to trace the evolutionary trajectories of the underlying genetic variants.

Three plumage color traits divergent across the wheatear genus are amenable to genetic mapping in the recently radiated *hispanica* complex, where they also evolved convergently (*25*), and are present as polymorphisms in three of four species (throat coloration) or are recombined by hybridization (mantle and neck side coloration) (*25*, *26*) (Fig. 1, A and B). The four species of the *hispanica* complex are distributed from western North Africa and Iberia to eastern China (Fig. 1, A and C): *Oenanthe hispanica*, *O. melanoleuca*, *O. cypriaca*, and *O. pleschanka* (*27*). Hybridization leads to phenotypic and genetic admixture (Fig. 1B) (*25*–*27*) in four regions: the western coast of the Black Sea, the Alborz mountains of northern Iran, the Caucasus, and the Mangyshlak peninsula in southwestern Kazakhstan ("Caspian," Fig. 1C) (*27*). These characteristics distinguish wheatears as an excellent system to investigate the genetic architecture and evolutionary history of plumage coloration.

Cryptic diversity, pervasive hybridization, and idiosyncratic ancestry composition of hybrid zones

To establish the geographic structure of genome-wide variation in the *hispanica* complex, we produced 379 phased genomes (assisted by linked-read information from 235 genomes) sampled over 9000 km across the complex's entire range, with dense sampling of hybrid zones (Fig. 1C and table S1).

Population genomic and phylogenomic analyses establish five ancestry groups that evolved in fast succession (Fig. 1, D to G, and figs. S1 and S2). Principal component analyses (PCAs) of genotype and haplotype frequencies distinguish separate clusters for the four known taxa and uncover a fifth cluster consisting of black-throated *O. pleschanka* phenotypes from the Pontic steppes west and north of the Black Sea (Fig. 1D and figs. S1 and S2), hereafter referred to as the Pontic lineage. The Pontic lineage is sister to *O. pleschanka* (Fig. 1E and fig. S3) but shows a distinct demographic trajectory (Fig. 1F).

Pairwise differentiation (F_{ST}) between lineages ranges from 11 to 41% (table S2). The distinctiveness of the five lineages is supported by (i) high coancestries within lineages in fineSTRUCTURE analyses (*28*) (figs. S4, and S5), (ii) the clear-cut clustering of the lineages based on coancestries (figs. S4 and S5), and (iii) nonlinear ancestry transitions across the three contact zones between *O. melanoleuca*, *O. pleschanka*, and the Pontic lineage (fig. S6). The short interval within which the inverse instantaneous coalescent rate (IICR) trajectories of the five lineages started diverging (Fig. 1F) together with cross-lineage coalescence times estimated from an ancestral recombination graph (ARG) (Fig. 1G) suggests that species split within a time window of about 60 thousand years (kyr) that occurred at the latest around 700 kyr ago (Fig. 1G; range of lowest to highest 95% confidence values, 678 to 727 kyr; table S3).

The *hispanica* complex's evolutionary history is shaped by hybridization. Aside from numerous individuals with intermediate ancestry compositions (Fig. 1D and fig. S4), D and f_4 statistics indicate abundant introgression (table S4). According to the taxa involved, each hybrid zone hosts a distinct ancestry composition, variably combining *O. melanoleuca*, *O. pleschanka*, and/or Pontic ancestries. The abundance of genetic and phenotypic admixture highlights pervasive hybridization in the *hispanica* complex and promises the power to map plumage coloration using individuals with admixed ancestries from different hybrid zones (Fig. 1, D and H, and figs. S2 and S4). All hybrid zones feature individuals with intermediate plumages (*26*, *27*), but individuals of mixed ancestry can be phenotypically indistinguishable from parental taxa (Fig. 1H and fig. S7). This combination of high and partly dissociated variation in plumage coloration and ancestry composition provides the power for the genetic mapping of color phenotypes divergent between parental species (*29*).

Modular architecture of plumage coloration

A transposable element insertion and protein-changing ASIP variation underpin throat coloration

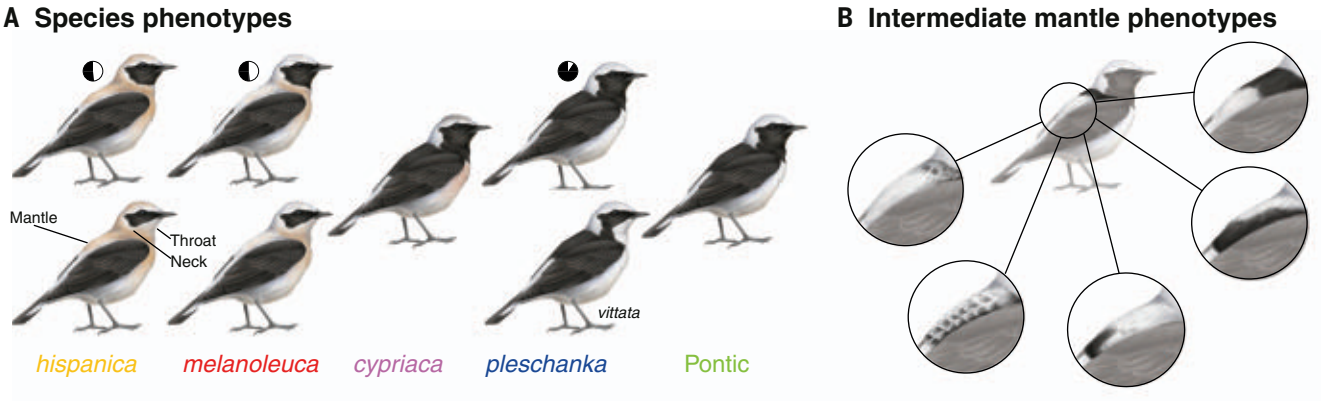
Throat coloration is polymorphic in three species of the *hispanica* complex (Fig. 1A). To investigate its genetic basis, we performed a genome-wide association study (GWAS) on 335 wheatears (57 white- and 278 black-throated) from all five lineages and hybrid zones based on linear mixed models that account for relatedness in GEMMA (*30*). The GWAS identified five single-nucleotide polymorphisms (SNPs) clustered within <15 kilo–base pairs (kbp) on chromosome 20 that are most strongly associated with throat coloration [negative logarithm of the odds ratio for linkage (–LOD) > 27; fig. S8A], all in the genomic region harboring a well-known gene within the melanogenesis pathway, the agouti signaling protein [ASIP; (*31*)] (Fig. 2, A and B, and fig. S9A; see supplementary text). The highest-scoring SNP together with the fully linked neighboring SNP ("throat SNPs") display nonsynonymous variation in consecutive codons in ASIP exon 2 that are otherwise highly conserved across birds (table S5). These throat SNPs are nearly completely associated with throat coloration across all five taxa and within *O. hispanica* and *O. melanoleuca* (Fig. 2C). The sequenced white-throated *pleschanka* ("*vittata*" phenotype, $n = 3$; Fig. 1A) are heterozygous for the same variants. Heterozygotes almost exclusively show white throat coloration, establishing the dominance of the white variant, similar to other species (*32*) and in line with ASIP expression preventing the differentiation or presence of melanocytes (*33*).

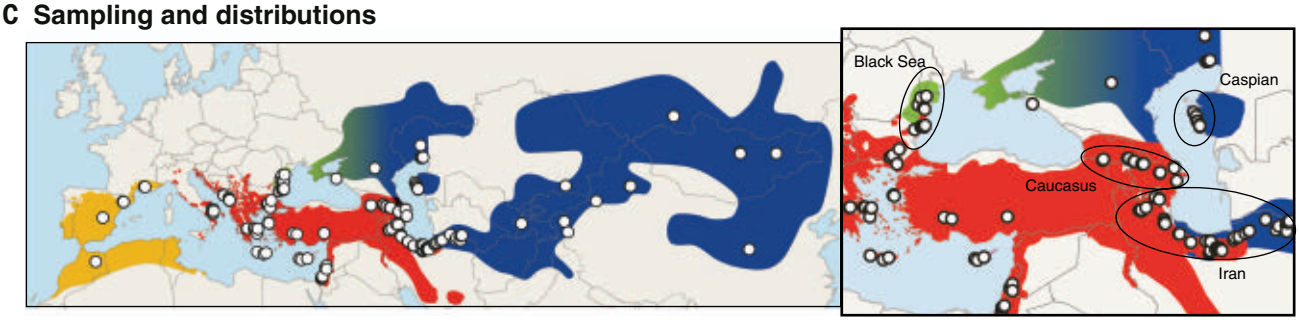
ASIP-related color variation is not usually underpinned by amino acid changes but rather with variation regulating ASIP expression [(*31*), but see (*34*)]. Therefore, we interrogated our data for linked variation hidden from typical GWASs, such as structural variants. A revised repeat annotation of chromosome 20 identified a >9-kbp-long, full-length long terminal repeat (LTR) retrotransposon ("throat LTR") in the reference genome (from a white-throated individual) 617 bp upstream of ASIP exon 2 (Fig. 2B). The presence of the throat LTR was nearly perfectly associated with white throat coloration in *O. hispanica* and *O. melanoleuca* (Fig. 2C).

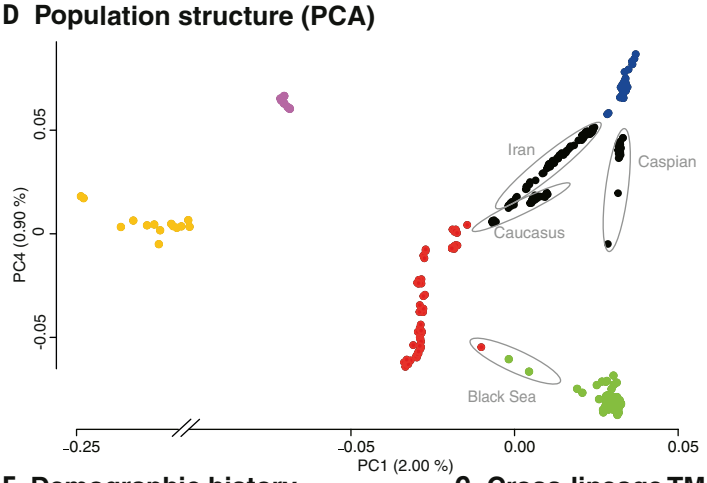
The interaction of throat SNP and LTR genotypes explains all throat color variation in *O. hispanica* and *O. melanoleuca* (associations of coloration with the remaining GWAS top hits were substantially weaker; table S6). From this, we infer that the throat LTR up-regulates ASIP expression, whereas the protein-changing throat SNP variation determines throat coloration (Fig. 2C). The LTR-SNP interaction is supported by rare recombinants for which the throat SNP and LTR genotypes do not match coloration ($n = 4$, genotypes marked by two asterisks in Fig. 2C). These mismatches suggest that the "white" ASIP protein is only expressed if an LTR up-regulates its expression (Fig. 2C, bottom two asterisks). Conversely, ASIP up-regulation by an LTR only produces white plumage if the up-regulated haplotype encodes the "white" ASIP protein (Fig. 2C, right two asterisks). Additional support for this model comes from ASIP haplotype structure. The model predicts that, to up-regulate expression of "white" protein, the LTR must collocate in cis on the same haplotype as the white-coding variant. In line with this prediction, for all heterozygotes with linked reads available from the ASIP region, linked-read barcodes exclusively ($n = 36$) or in majority ($n = 2$) placed the throat LTR and the white throat variant on the same haplotype (table S7).

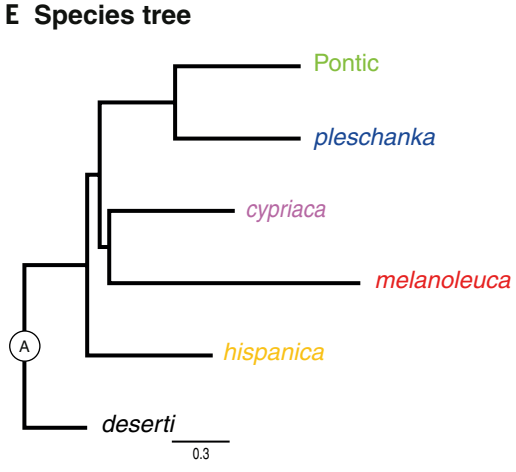
ASIP expression may be regulated differently in the white-throated *vittata* phenotype of *O. pleschanka*. Although the rarity of this phenotype and the accordingly limited data we had at hand prevents us from making firm conclusions, several lines of evidence point in this direction. We found no evidence for the LTR's presence in any black- or white-throated (*vittata*) individuals of *O. pleschanka* and the Pontic lineage. For the three (white-throated) *vittata* individuals, neither a linked-read–based de novo assembly (fig. S10 and data S1)

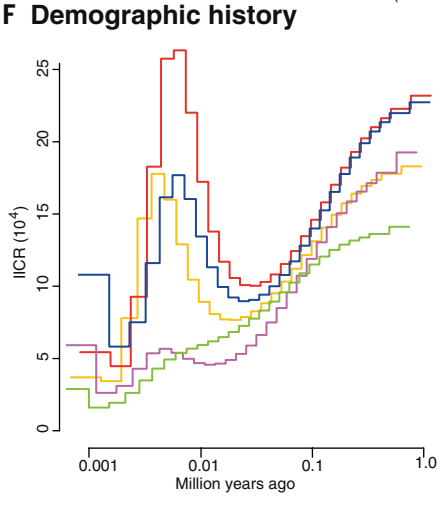
[1]Division of Evolutionary Ecology, Institute of Ecology and Evolution, University of Bern, Bern, Switzerland. [2]Swiss Ornithological Institute, Sempach, Switzerland. [3]Department of Population Ecology, Institute of Ecology and Evolution, Friedrich-Schiller-University Jena, Jena, Germany. [4]Department of Genetics and Bioinformatics, Swedish Natural History Museum, Stockholm, Sweden. [5]Department of Zoology, Stuttgart State Museum of Natural History, Stuttgart, Germany. [6]Department of Biological Sciences, University of Cyprus, Nicosia, Cyprus. [7]Département d'Ecologie et Evolution, Université de Lausanne, Lausanne, Switzerland. [8]Section of Ecology and Systematics, Department of Biology, National and Kapodistrian University of Athens, Athens, Greece. [9]Department of Bioecology, Baku State University, Baku, Azerbaijan. [10]Dokuz Eylül University, Science Faculty, Department of Biology, İzmir, Türkiye. [11]Department of Biology & Zoological Innovations Research Department, Institute of Applied Zoology, Faculty of Sciences, Ferdowsi University of Mashhad, Mashhad, Iran. [12]TYTO – Association for the Management and Conservation of Biodiversity in Agricultural Ecosystems, Larisa, Greece. [13]Pafos Zoo, Paphos, Cyprus. [14]Biblioteca de Ciència i Tecnologia, Universitat Autònoma de Barcelona, Bellaterra, Spain. [15]Studio Naturalistico MILVUS, Altamura, Italy. [16]Instituto de Investigación en Recursos Cinegéticos (IREC), CSIC-UCLM-JCCM, Ciudad Real, Spain. [17]Center for Wildlife Disease Ecology, Institute of Ecology, Ilia State University, Tbilisi, Georgia. [18]Attica Zoological Park, Athens, Greece. [19]Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, China. [20]University of Chinese Academy of Sciences, Beijing, China. [21]State Key Laboratory of Biocontrol, School of Ecology, Sun Yat-sen University, Shenzhen, China. [22]Centre for Biological Diversity, University of St Andrews, St Andrews, United Kingdom. [23]Department of Parasitology and Zoology, University of Veterinary Medicine, Budapest, Hungary. [24]Department of Forest Sciences, Faculty of Agriculture and Forestry, University of Helsinki, Helsinki, Finland. [25]STAR-UBB Institute, Babes-Bolyai University, Cluj-Napoca, Romania. [26]HUN-REN-UVMB Climate Change: New Blood-Sucking Parasites and Vector-Borne Pathogens Research Group, Budapest, Hungary. [27]Independent researcher, Hauteville, Switzerland. [28]National Museum of Natural History Sofia, Bulgarian Academy of Sciences, Sofia, Bulgaria. [29]Association for the Conservation of Biodiversity of Kazakhstan (ACBK), Almaty, Kazakhstan. [30]Independent researcher, Cirkulane, Slovenia. [31]Department of Zoology and Physiology, Baku State University, Baku, Azerbaijan. [32]Applied Zoology Center, Institute of Zoology, Ministry of Science and Education, Baku, Azerbaijan. [33]Science for Life Laboratory, Division of Gene Technology, KTH Royal Institute of Technology, Stockholm, Sweden. [34]Science for Life Laboratory, Department of Biosciences and Nutrition, Karolinska Institute, Stockholm, Sweden. [35]Centre for Molecular Biodiversity Research, Leibniz Institute for the Analysis of Biodiversity Change, Zoologisches Forschungsmuseum A. Koenig, Bonn, Germany. [36]Bonn Institute for Organismic Biology (BIOB) – Animal Biodiversity, University of Bonn, Bonn, Germany. [37]Department of Organismal Biology – Systematic Biology, Uppsala University, Uppsala, Sweden. [38]ISEM, Université Montpellier, CNRS, EPHE, IRD, Montpellier, France. [39]Natural History Museum Bern, Bern, Switzerland. [40]Division of Population Genetics, Institute of Ecology and Evolution, University of Bern, Bern, Switzerland. [41]German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Leipzig, Germany. ***Corresponding author. Email: reto.burri@vogelwarte.ch** †Deceased. ‡Present address: Olink Proteomics, part of Thermo Fisher Scientific, Uppsala, Sweden.

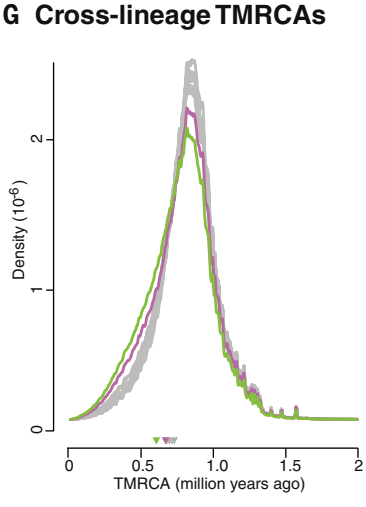


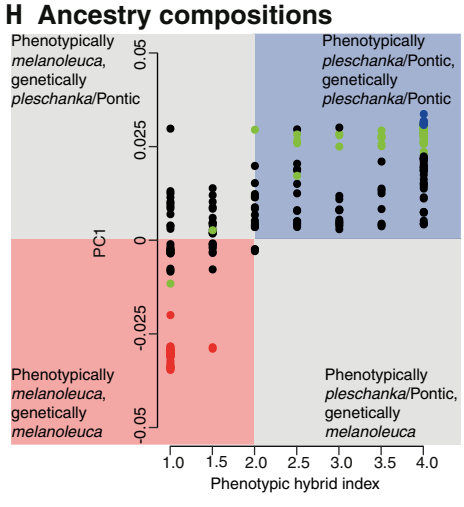












BIRD ARTWORK: M. SCHWEIZER.

Fig. 1. Plumage color variation and evolutionary history of the *hispanica* complex. (**A**) Color phenotypes and frequencies of throat phenotypes. Mantle and neck sides are white in *O. hispanica* and *O. melanoleuca* and black in the remaining taxa. *O. hispanica*, *O. melanoleuca*, and *O. pleschanka* display a throat color polymorphism [pies indicate species-wide frequencies estimated from (*98*)]. (**B**) Hybrid-type mantle color variation. Black coloration on hybrids' mantles can vary from absent, overmottled, and reduced black to fully black. (**C**) Species ranges and sampling. Colors follow those in (A). Samples are depicted by white dots. Cyprus is shown magnified. The right-hand box zooms onto hybrid zones' locations (encircled). Species distributions are adapted from maps provided by Birdlife International (*99*). (**D**) Population structure inferred by PCA. For further PC axes, see fig. S1. Black, encircled dots depict individuals from hybrid zones. (**E**) Species tree inferred using the multispecies coalescent. The node for which ancestral states of genetic variation underpinning color phenotypes were estimated is highlighted (circled A). All nodes have 100% local posterior probability. (**F**) Demographic history of the lineages. (**G**) Genome-wide distributions of TMRCAs between lineages show that lineages split within a narrow time window. Most distributions are similar (gray); those between *O. melanoleuca* and *O. cypriaca* (purple) and those between Pontic and *O. pleschanka* (green) show more recent TMRCAs. Tick marks indicate mean TMRCAs between lineages (confidence intervals too small to be visible; table S3). (**H**) Highly variable ancestry compositions of color phenotypes provide high power for GWASs. Shown is individuals' ancestry, as estimated from PCA (low values, *O. melanoleuca*; high values, *O. pleschanka*), in relation to phenotypic hybrid index (1, white mantle and neck sides; 4, black mantle and neck sides). Additional PC axes are shown in fig. S7. Colored points follow those in (A).

nor short-read–based LTR genotyping indicated the throat LTR's presence, despite the white throat SNP variant being present. Together with the observation that *vittata* displays a white throat despite featuring the same recombinant haplotype as two black-throated *O. melanoleuca* individuals (Fig. 2C, bottom), this suggests that unknown variation may up-regulate ASIP expression in the *vittata* phenotype of *O. pleschanka*.

Additive ASIP-based architecture of mantle and neck side coloration

Mantle and neck side coloration are divergent between white-mantled *O. melanoleuca* and *O. hispanica* and black-mantled *O. pleschanka*, Pontic, and *O. cypriaca*, with the same color variants shared by non-sister species (Figs. 1, A and E).

To infer the genetic variation underpinning the species-specific mantle and neck side coloration, we leveraged information from 230 individuals from all four hybrid zones (Fig. 1C) in which mantle coloration is recombined into diverse genomic backgrounds (Fig. 1H and fig. S7). GWAS identified 17 variants (14 SNPs and 3 indels, hereafter referred to as "mantle loci;" Fig. 2A and fig. S9, B and C; see supplementary text) strongly associated with mantle coloration (–LOD > 15; fig. S8B), of which 15 are located within <45 kbp upstream of ASIP exon 2, and 2, in intron 2 (Fig. 2B).

Bayesian linear regressions support an additive architecture of mantle coloration. Individuals with a higher number of white variants at mantle loci show increasingly white mantles (Fig. 2D and table S8). Entirely black mantles are expressed with highest probability only in individuals with fewer than 10 white alleles (out of 34 alleles or 17 mantle loci) (Fig. 2D). This suggests partial dominance of white mantle coloration.

The GWAS for neck side coloration identified overall weaker associations (Fig. 2A). The overlap of the identified SNPs with mantle loci (Fig. 2B and fig. S9C; see supplementary text) suggests that neck side coloration is an extension of mantle coloration associated with the same genetic variation. The observation that birds with black neck sides but white mantles are the most frequent hybrid phenotypes leads us to hypothesize that the penetrance of white is lower for neck sides than for the mantle, which may result from embryonic prepatterning by ASIP (*35*). In line with this, more white variation would be required at mantle loci to result in white neck sides than are required to result in a white mantle (see shifted intersection between Fig. 2, D and E). Henceforth, we focus on mantle coloration, which implicitly includes neck side coloration.

ASIP expression and isoform structures corroborate the genetic architecture of wheatear plumage coloration

To corroborate the statistically inferred genetic architectures of throat and mantle coloration, we investigated gene expression patterns in regrowing feathers. To this end, we chose *O. melanoleuca* and *O. cypriaca* as representatives of alternative colorations and performed single-molecule real-time isoform sequencing (Iso-Seq) of RNA for feather pools representing alternatively black or white coloration of throat and mantle, and, in addition, feathers from the shoulders and belly that are homogeneously black (shoulders) or white (belly) across species (Fig. 3A).

ASIP expression profiles of regrowing feathers corroborate ASIP's role in regulating wheatear plumage coloration. ASIP expression is about four times higher in white than in black feathers for the major isoform ($t = 3.6$, $P = 0.007$; adjusted coefficient of determination $R^2 = 0.57$; Fig. 3B; see supplementary text) also when accounting for the feathers' species of origin (color, $t = 3.9$, $P = 0.006$; species, $t = -1.5$, $P = 0.19$; adjusted $R^2 = 0.62$). This result holds when jointly analyzing all detected isoforms ($t = 4.6$, $P = 0.002$; adjusted $R^2 = 0.69$; accounting for species: color, $t = 5.6$, $P < 10^{-3}$; species, $t = -2.0$, $P = 0.08$; adjusted $R^2 = 0.78$; fig. S11). The higher expression in white than in black feathers is in line with ASIP's role in preventing melanin synthesis when acting early during feather growth (*33*).

Moreover, ASIP isoform structures confirm the role of the GWAS hits in ASIP regulation. Eight of the 17 mantle loci, including the top GWAS hits, are situated within or immediately upstream of the three alternative forms of exon 1 (fig. S12). This provides evidence for these genomic regions' involvement in ASIP regulation that is entirely independent of the statistical associations established by GWAS and establishes mantle loci as likely key regulators of ASIP expression in mantle feathers. For white throat color, all transcription start sites are situated further 5′ of ASIP exon 2 than the throat LTR. Together with the reverse complementary orientation of the LTR compared with the ASIP gene (Fig. 2B), this suggests that the LTR is unlikely to provide a transcription start site. Rather, the throat LTR acts as a (throat-specific) enhancer of ASIP expression, as documented for transposable element (TE) insertions upstream of ASIP in other species (*36*).

The patterns of gene expression and isoform variation recovered from regrowing feathers thus provide compelling support for the ASIP gene's key role in regulating melanin-based coloration in wheatears.

Introgression of throat coloration

Throat coloration is polymorphic in three species, with intermediate frequencies of the black and white phenotypes in *O. hispanica* and *O. melanoleuca* and with the black phenotype predominant (>98%) in *O. pleschanka* (*27*) (Fig. 1A). It was unclear whether this polymorphism arose from incomplete lineage sorting of ancestral variation or from introgression of variation that evolved following the species' divergence.

We established that the throat color polymorphism evolved by introgression of white ASIP variation among the throat-polymorphic species. Comparisons of times to the most recent common ancestor (TMRCAs) inferred in tsinfer (*37*) and tsdate (*38*) show that (i) the polymorphism evolved only after species diverged, with (ii) white being the derived state, and that (iii) the ASIP gene body shows distinctive footprints of introgression, with a divergence of the white ASIP haplotypes between *O. hispanica* and *O. melanoleuca* dating this introgression to 403 kyr ago, after *O. melanoleuca* split from *O. cypriaca* (Fig. 4A and tables S3 and S9; see supplementary text).

Next, we established that the white-throated phenotype originated in *O. melanoleuca*. Population genetic expectations are that (i) the species in which the white ASIP haplotype originated displays higher

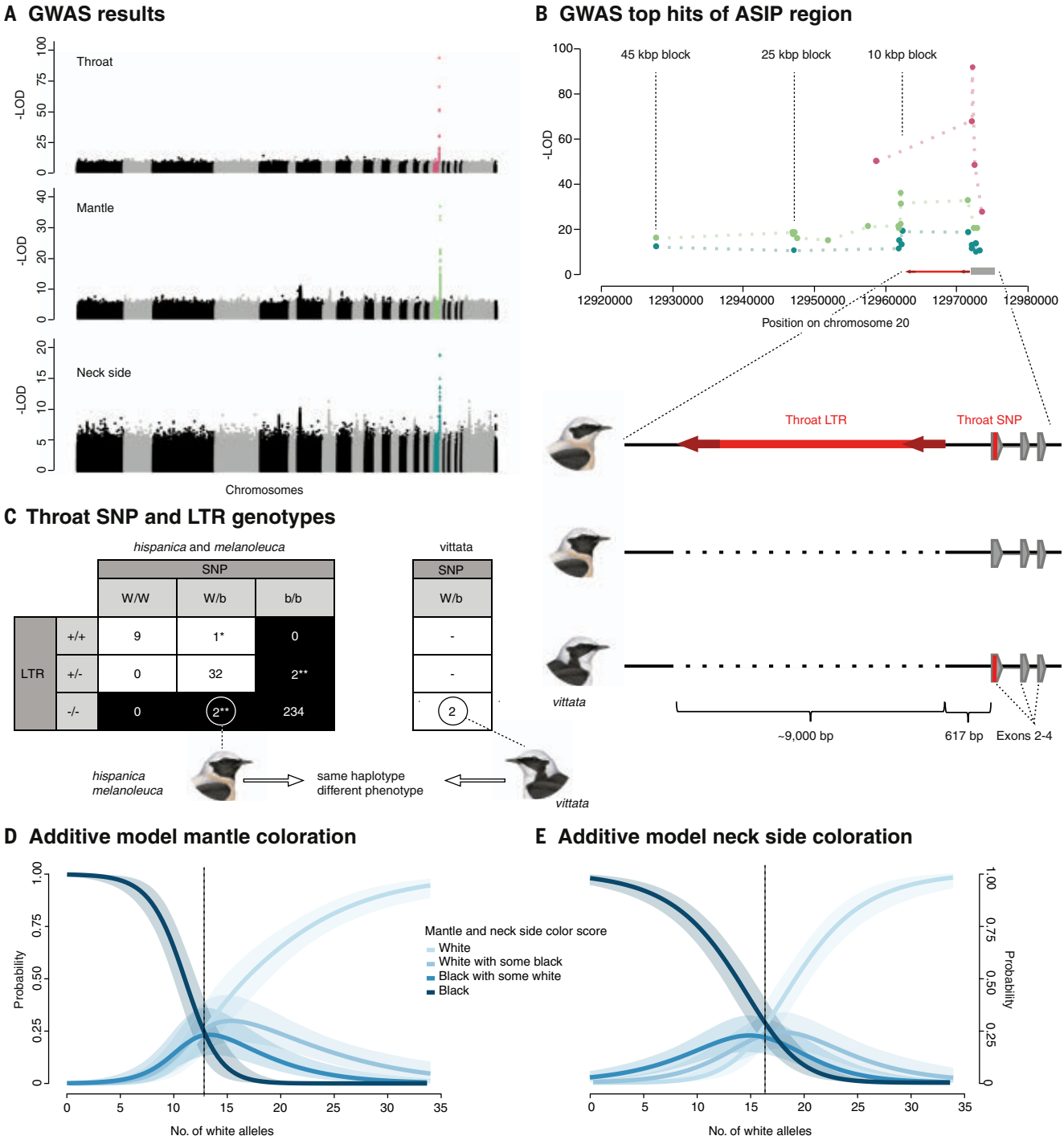


Fig. 2. Genetic architectures of plumage coloration. (**A**) GWAS results across the genome. Chromosomes are marked in black and gray; chromosome 20 is colored. (**B**) Distribution of GWAS top SNPs for the three color phenotypes within the ASIP region [colors follow those in (A)] and schematics of the haplotypes underlying throat color phenotypes in *O. melanoleuca* (top two) and the *vittata* phenotype of *O. pleschanka* (bottom). Arrowed boxes depict annotated ASIP exons. The throat LTR and SNP are highlighted in red. Absence of the throat LTR is depicted by dotted lines. (**C**) Genotypes of the throat SNPs and throat LTR. Boxes are shaded according to throat coloration. Recombinants in *O. hispanica* and *O. melanoleuca* are highlighted by asterisks: Recombinants supporting the interaction of the throat LTR and SNP in determining coloration are indicated with two asterisks, and recombinants not informative in this respect are indicated with one asterisk. All genotypes agree with the phenotypes predicted by a model in which the presence of the throat LTR (+/+ or +/−) up-regulates ASIP expression and, in the presence of a white throat SNP variant (W/W or W/b), confers white throat coloration, whereas an absent throat LTR (−/−) or homozygous black throat SNP (b/b) confers black throat coloration. See also table S6 for matches of observed with predicted phenotypes of the throat SNPs and lower-ranking GWAS hits. (**D**) Bayesian linear regressions of mantle variation. Shading shows 95% confidence intervals. The probability of observing a black mantle decreases, whereas the probability of observing a white mantle increases with an increasing number of white alleles (out of 34 possible at the 17 mantle loci) observed at mantle loci. Intermediate mantle colorations are most likely observed at intermediate numbers of white alleles. (**E**) Bayesian linear regressions of neck side variation. The probability of observing a black neck decreases, whereas the probability of observing a white neck increases with an increasing number of white alleles observed at mantle loci. Intermediate neck colorations are most likely observed at intermediate numbers of white alleles.

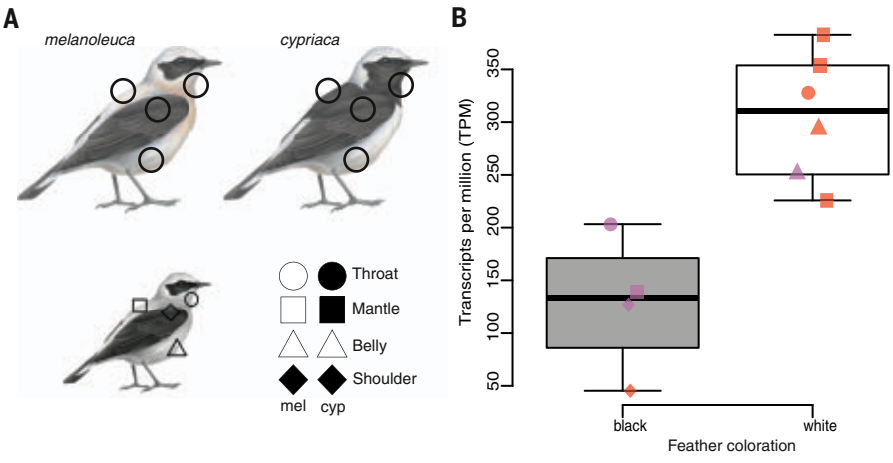


Fig. 3. Association of ASIP expression with feather coloration. (**A**) Design of the Iso-Seq experiment. Iso-Seq was performed once on each combination of feather tract, coloration, and species, except for white mantle, for which three pools from *O. melanoleuca* were sequenced (tables S16 and S17). mel, *melanoleuca*; cyp, *cypriaca*. (**B**) Association of the expression of the major ASIP isoform with feather coloration (TPM, transcripts per million reads). Symbols show individual values and follow those in (A); colors indicate species (red, *O. melanoleuca*; purple, *O. cypriaca*). ASIP expression is higher in white than in black feathers ($t = 3.6$, $P = 0.007$; adjusted R^2=0.57).

diversity (thus, has older TMRCAs) of the derived white ASIP haplotype, whereas, conversely, (ii) introgression increases both the diversity of black ASIP haplotypes and of overall ASIP diversity in the introgressed species. In line with an origin of the white haplotype in *O. melanoleuca*, we found an older TMRCA among white ASIP haplotypes in *O. melanoleuca* (395 kyr) than in *O. hispanica* (320 kyr; Fig. 4A, label 6, and table S9) and older TMRCAs among black ASIP haplotypes and across all ASIP haplotypes in *O. hispanica* (black haplotypes, 564 kyr; all haplotypes, 535 kyr) than in *O. melanoleuca* (black haplotypes, 554 kyr; all haplotypes, 510 kyr; table S10). Despite the confidence intervals of the white haplotypes' TMRCA overlapping between *O. hispanica* and *O. melanoleuca*, this result provides conclusive evidence for an origin of the derived white throat phenotype in *O. melanoleuca* (see supplementary text; fig. S13 and table S10). The white throat phenotype evolved through new protein-coding variation and insertion of the throat LTR in *O. melanoleuca* before being introgressed into *O. hispanica*.

The origin of the white-throated *vittata* phenotype in *O. pleschanka* was likely more intricate. For *O. pleschanka*, owing to the small sample size of white haplotypes ($n = 3$) and their recombined structure, we could not obtain the same level of evidence for introgression as observed in *O. melanoleuca*. Nonetheless, an introgressed origin, similar to that in *O. hispanica*, seems plausible, given well-documented hybridization with *O. melanoleuca* (*26*, *27*) (Fig. 1D and table S4). The absence of the throat LTR in *vittata* (fig. S10 and data S1) suggests that homologous recombination may have dissociated the white coding variant from the throat LTR, as also observed in the five recombinants marked with asterisks in Fig. 2C. In the absence of the throat LTR, it seems reasonable to assume that in *vittata* throat feathers, ASIP expression is up-regulated by undetected new genetic variation.

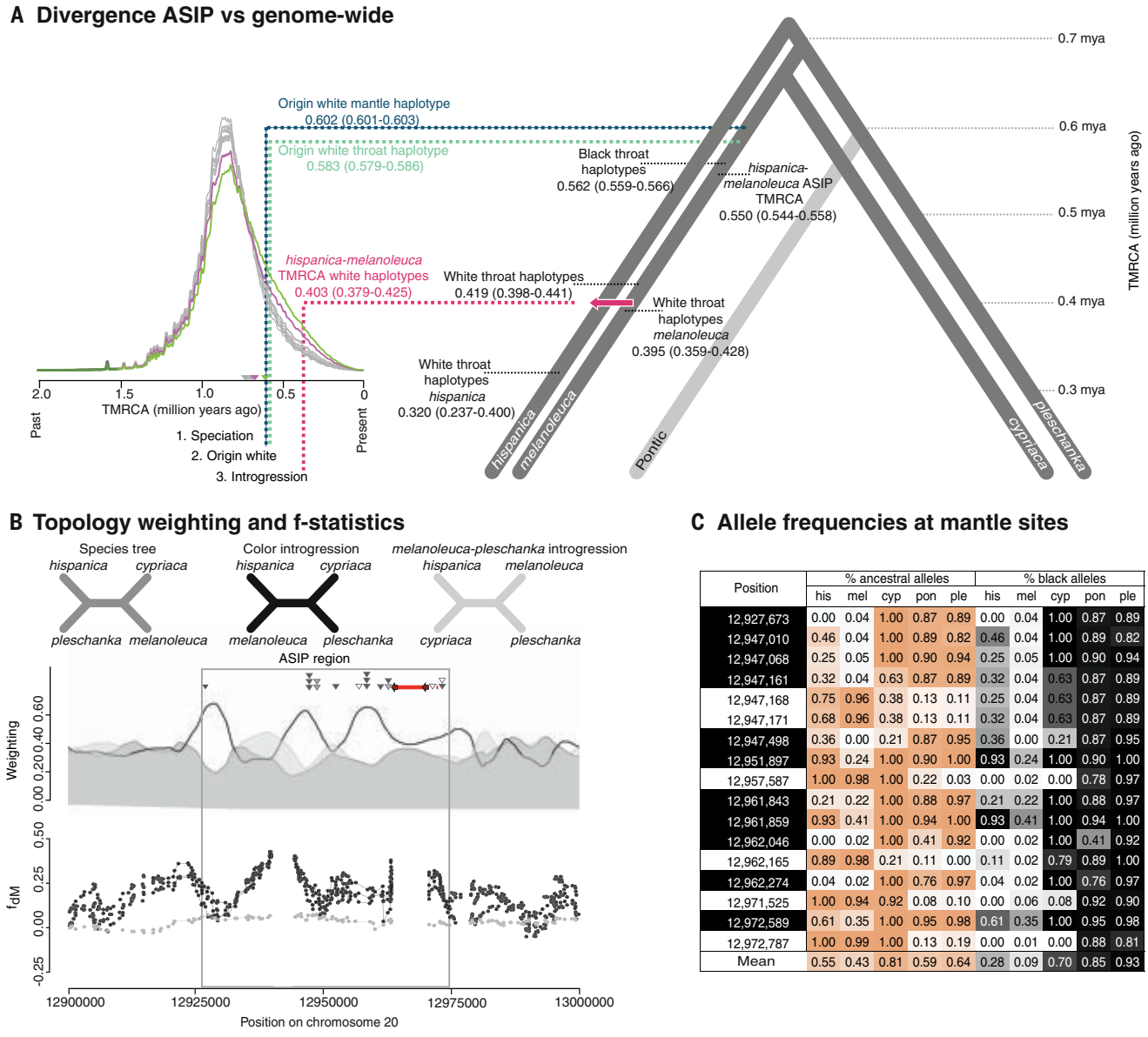
The successful introgression and maintenance of the throat color polymorphism suggest a role of throat coloration in wheatear ecology. Assortative and disassortative mating are unlikely to explain the polymorphism's maintenance: Contrary to expectations under such a scenario, we do not observe an excess of homo- or heterozygotes at the throat SNPs (table S11). Moreover, the two throat phenotypes co-occur on small spatial scales, and no differences in habitat or resource use are known despite rather well-studied species ecology (*27*). We therefore tested the hypothesis that the two throat phenotypes occupy subtly different microniches with different nutrient sources. To this end, we first analyzed triple-element (nitrogen, carbon, and sulfur) stable isotope composition in the whole blood of 161 males of *O. melanoleuca* ($n_{black} = 49$, $n_{white} = 43$) and Iranian hybrid ancestries ($n_{black} = 56$, $n_{white} = 13$). Blood isotope composition provides an integrated isotopic signature over several weeks, reflecting dietary intake and ecological interactions during this period. It can indicate differences in trophic niche [nitrogen, δ^{15}N; (*39*)]; habitat composition [carbon, δ^{13}C; (*40*)]; and foraging niche, dietary source, and habitat use [sulfur, δ^{34}S; (*41*)]. The alternative throat phenotypes showed no significant differences in δ^{13}C and δ^{34}S values (fig. S14). However, irrespective of whether analyzed for all individuals or *O. melanoleuca*, Greek, or Iranian males separately (table S12), white-throated males exhibited a narrower distribution as well as, on average, lower δ^{15}N values than black-throated individuals (Fig. 5 and fig. S15). This result persisted after accounting for individual and geographic correlates (table S12) and for unequal sample sizes of the two throat phenotypes (fig. S16). Second, we were interested in whether the same association persisted in *O. hispanica*, where the white throat phenotype introgressed from *O. melanoleuca*. Despite limited sample sizes ($n_{black} = 15$, $n_{white} = 8$), this analysis shows that white throats are associated with lower δ^{15}N values in *O. hispanica*, too (Fig. 5, fig. S16, and table S12). These results suggest that white-throated individuals prey on insects from a narrower and lower trophic level than black-throated individuals.

On the basis of these results, we speculate that the throat polymorphism is related to the trophic niche of prey and thus the wheatears' foraging strategy, such as observed in other birds: In barn owls, white underbody coloration aids hunting by prolonging prey freezing times in moonlight (*42*), whereas Nearctic warblers and redstarts display white feather tracts to flush prey and enhance foraging success (*43*, *44*). White-throated wheatears may be better at exploiting a specialized niche in a related manner, but further research is required to infer the adaptive role of the throat polymorphism. Alternatively, black-throated birds may be more aggressive (*45*) and exclude white-throated conspecifics from part of the species' niche. However, we deem such a scenario less likely, as it involves a lower fitness of the white throat phenotype and does not explain the polymorphism's maintenance and successful introgression. Therefore, and because the difference in nitrogen isotope composition between throat phenotype seems strongest in *O. hispanica*, the species into which white throat coloration successfully introgressed to high frequencies (Fig. 1A), we favor a scenario of niche specialization of the white throat phenotype. This would suggest that the recruitment of the white throat haplotype from *O. melanoleuca* into *O. hispanica* and *O. pleschanka* constitutes an event of likely adaptive introgression that enabled the exploitation of a specialized foraging niche.

Mantle color evolution by cointrogression of new variation

Mantle coloration is more similar between more distantly related lineages than between sister lineages of the *hispanica* complex (Fig. 1, A and E). We show that this pattern is rooted in the evolution of white mantle coloration in *O. melanoleuca* followed by its cointrogression with white throat coloration into *O. hispanica*.

Our results indicate that, originally, all members of the *O. hispanica* complex displayed black mantle plumage. Ancestral state reconstructions show that black variants were ancestral at 11 mantle loci at the node at which the *O. hispanica* complex split from *O. deserti* (node A, Fig. 1E). Furthermore, except *O. melanoleuca*, all species show moderate



Position	% ancestral alleles					% black alleles				
	his	mel	cyp	pon	ple	his	mel	cyp	pon	ple
12,927,673	0.00	0.04	1.00	0.87	0.89	0.00	0.04	1.00	0.87	0.89
12,947,010	0.46	0.04	1.00	0.89	0.82	0.46	0.04	1.00	0.89	0.82
12,947,068	0.25	0.05	1.00	0.90	0.94	0.25	0.05	1.00	0.90	0.94
12,947,161	0.32	0.04	0.63	0.87	0.89	0.32	0.04	0.63	0.87	0.89
12,947,168	0.75	0.96	0.38	0.13	0.11	0.25	0.04	0.63	0.87	0.89
12,947,171	0.68	0.96	0.38	0.13	0.11	0.32	0.04	0.63	0.87	0.89
12,947,498	0.36	0.00	0.21	0.87	0.95	0.36	0.00	0.21	0.87	0.95
12,951,897	0.93	0.24	1.00	0.90	1.00	0.93	0.24	1.00	0.90	1.00
12,957,587	1.00	0.98	1.00	0.22	0.03	0.00	0.02	0.00	0.78	0.97
12,961,843	0.21	0.22	1.00	0.88	0.97	0.21	0.22	1.00	0.88	0.97
12,961,859	0.93	0.41	1.00	0.94	1.00	0.93	0.41	1.00	0.94	1.00
12,962,046	0.00	0.02	1.00	0.41	0.92	0.00	0.02	1.00	0.41	0.92
12,962,165	0.89	0.98	0.21	0.11	0.00	0.11	0.02	0.79	0.89	1.00
12,962,274	0.04	0.02	1.00	0.76	0.97	0.04	0.02	1.00	0.76	0.97
12,971,525	1.00	0.94	0.92	0.08	0.10	0.00	0.06	0.08	0.92	0.90
12,972,589	0.61	0.35	1.00	0.95	0.98	0.61	0.35	1.00	0.95	0.98
12,972,787	1.00	0.99	1.00	0.13	0.19	0.00	0.01	0.00	0.88	0.81
Mean	0.55	0.43	0.81	0.59	0.64	0.28	0.09	0.70	0.85	0.93

Fig. 4. Introgressed origins of ASIP variation. (**A**) ARG-based reconstruction of of ASIP haplotypes' evolutionary history within the species tree (right) and distributions of interspecific TMRCAs (left). White haplotypes evolved after relevant species splits (blue and green dotted lines). The *O. hispanica–O. melanoleuca* ASIP TMRCA (upper right black dotted line) is more recent than these species' genome-wide TMRCA, indicating ASIP introgression. The TMRCA among white haplotypes is more recent than that among black haplotypes (upper and middle left black dotted lines) and establishes the white haplotype as derived. The TMRCA of white haplotypes in *O. melanoleuca* was older than that of white haplotypes in *O. hispanica* (lowest right and left black dotted lines) and establishes white coloration's origin in *O. melanoleuca*. The TMRCA between white haplotypes in *O. hispanica* and *O. melanoleuca* (pink dotted line) pinpoints the time of ASIP introgression. TMRCAs' confidence intervals are too small to be visualized (fig. S13 and tables S3 and S9). (**B**) Topology weighting with topologies and their interpretation (top) and f_{dM} tests for *O. melanoleuca–O. hispanica* introgression (bottom) variably including one of three black-mantled lineages (black) or exclusively black-mantled lineages (gray). Triangles highlight mantle loci that ancestrally featured black variation (black), polymorphic variation (gray), or white variation (white). Red blocks indicate the throat LTR and throat SNP. Regions harboring color loci show an excess of the introgression topology and footprints of *O. melanoleuca–O. hispanica* introgression (positive f_{dM}). The discrepancy between high f_{dM} in the comparison including *O. melanoleuca* (black) and ones only including black-mantled lineages (gray) consolidates evidence for introgression. (**C**) Frequencies of ancestral and black alleles at mantle loci. Darkening shades indicate increasing allele frequencies. Positions' colors indicate ancestral states. The high frequencies of derived variants and lowest frequencies of black alleles in *O. melanoleuca* and higher frequencies of ancestral black alleles in *O. hispanica* demonstrate the origin of white mantle variation in *O. melanoleuca* and introgression of the white variation into an originally black-mantled background in *O. hispanica*.

(*O. hispanica*) to high frequencies (all black-mantled species) of black variants at most mantle loci (Fig. 4C). The considerable frequencies of black variants at three ancestrally white sites in all species but *O. melanoleuca* (Fig. 4C) suggest that these were polymorphic in the ancestor of the complex. Given the additive genetic architecture (Fig. 2D), the ancestral populations would have likely displayed black mantle coloration.

Then, *O. melanoleuca* evolved derived white variation at the 11 ancestrally black mantle loci, and white variation introgressed into *O. hispanica*. *O. melanoleuca* features white variation at by far the

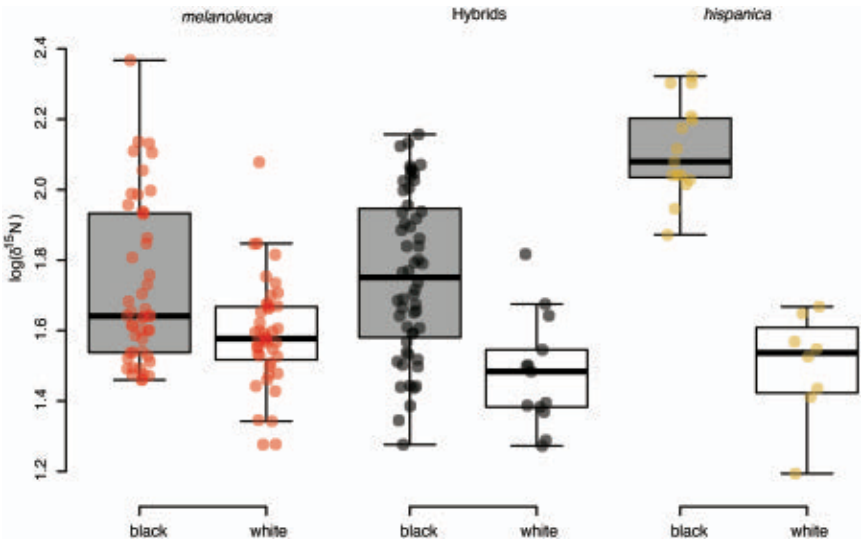


Fig. 5. Association of nitrogen stable isotopes with throat phenotype. Results are shown for analyses including all *O. melanoleuca*, hybrids, and *O. hispanica*. Points show individual values for *O. melanoleuca* (red), hybrid (black), and *O. hispanica* (yellow). All comparisons are significant (*O. melanoleuca*, $t = -3.2$, $P = 0.002$; hybrids, $t = -4.1$, $P < 10^{-3}$; *hispanica*, $t = -8.3$, $P < 10^{-5}$; table S12). Additional comparisons including Greek *O. melanoleuca* separately, all *O. melanoleuca* and hybrids together, and all individuals together are shown in fig. S15.

highest frequencies (91 versus 72% in *O. hispanica*; Fig. 4C). The considerable frequencies of black variation in *O. hispanica* (Fig. 4C) suggests that this species was originally black-mantled but largely swamped by white *O. melanoleuca* variation. This scenario is supported by the same evidence as introgression of throat variation. The genomic regions underpinning mantle coloration display predominantly topologies related to mantle coloration (Fig. 4B, figs. S17 and S18, and table S4); the TMRCAs between black and white haplotypes upstream of ASIP are younger than TMRCAs between species, and the white haplotypes' coalescence between *O. melanoleuca* and *O. hispanica* is younger than these species' split (Fig. 4, figs. S13 and S17, and tables S3, S9, and S10).

Several lines of evidence suggest an involvement of selection in mantle color evolution, especially in *O. melanoleuca*. Compared with the site frequency spectra (SFS) polarized according to ancestry, in all species but *O. cypriaca*, SFS polarized according to coloration show shifts toward the species' coloration (Fig. 6A). The strongest shifts are observed in *O. melanoleuca*, and conspicuous haplotype structures and TMRCAs corroborate an involvement of selection in driving white variation toward fixation, especially upstream of the most distal mantle loci (Fig. 6B, label 1, and fig. S19). Species other than *O. melanoleuca* do not feature comparable footprints of positive selection (Fig. 6B and fig. S19), likely because of a predominant involvement of ancestral variation, and, in *O. hispanica*, introgression.

We conclude that white mantle coloration evolved adaptively in *O. melanoleuca* and introgressed into *O. hispanica* along with the linked white throat variation.

Multistage introgression underpins color convergence and polymorphism

Within the *hispanica* complex, the evolution of species' plumages as we know them today predominantly roots in past introgression. Our results suggest that originally all members of the *hispanica* complex displayed black throat and mantle coloration before white mantle coloration evolved in *O. melanoleuca* through mutations upstream of ASIP that were fixed by positive selection (Fig. 7A). On this white-mantled genomic background, new protein-coding ASIP variation together with an LTR retrotransposon insertion upstream of ASIP led to the emergence of the white-throated phenotype in *O. melanoleuca* (Figs. 4A and 7A). White throat coloration then introgressed into *O. hispanica*, where, as in *O. melanoleuca,* it is maintained as a balanced polymorphism. Possibly owing to its linkage to white throat variation, white mantle variation cointrogressed into *O. hispanica*, where, due to its additive architecture with a partial dominance of white, it confers white mantle coloration despite still segregating black mantle variation (Fig. 4C). Footprints of positive selection (Fig. 6B, label 1), though much subtler than in *O. melanoleuca*, suggest that mantle color introgression might also have been supported by selection in *O. hispanica* (Fig. 6B). In *O. pleschanka*, hybridization with *O. melanoleuca* resulted in the introgression of white coding variation at ASIP, but not cointrogression of variation 5′ of ASIP. We interpret this in terms of selection against introgression of white mantle coloration; nearly fixed derived black variation at three of the mantle loci in *O. pleschanka* and the Pontic lineage (but not *O. cypriaca*) along with slightly longer–than–average haplotypes in the region flanked by two of these sites suggest that black mantle coloration is under positive selection in *O. pleschanka* and the Pontic lineage (Figs. 4C and 6B, label 2). For white throat variation, only the throat variant successfully introgressed into *O. pleschanka*. Introgression thus only planted the seed for the white-throated *vittata* phenotype. It thus appears that introgression shuffled plumage coloration among wheatear species and led to the convergent evolution of throat and mantle color phenotypes, similar to the reshuffling of morphologic or color traits documented from cichlids, *Heliconius* butterflies, Capuchino seedeaters, and Darwin's finches (*3*–*5*, *13*, *14*).

Introgression further contributed to the formation of intermediate phenotypes observed in extant hybrid zones (Figs. 7, B and C), where hybridization and backcrossing recombine ASIP variation into recombined haplotypes and color phenotypes (Figs. 1B and 7, B and C, and data S2). Especially Iranian and Caucasian individuals of mixed ancestries highlight how the fine-scale reshuffling of mantle and throat variation into recombined ASIP haplotypes directs the expression of intermediate phenotypes (Fig. 7C and data S2). Different from these recombined phenotypes of genome-wide admixed individuals, the Pontic and Caspian populations feature white-mantled to *O. melanoleuca*–type (white mantle with white neck sides) phenotypes in individuals with entirely *O. pleschanka* or Pontic ancestries (Fig. 7). This mismatch between phenotype and genome-wide ancestry takes its extreme in an *O. melanoleuca* phenotype individual from the Caspian population whose genome-wide ancestry exactly matches those of *O. pleschanka* phenotypes from the same region (no. 11, Fig. 7). Like similar Pontic and Caspian individuals, this individual displays ASIP haplotypes matching its coloration. This suggests that introgression is restricted to a small proportion of the genome. Accordingly, compared with Pontic and Caspian *O. pleschanka*–type phenotypes, *O. melanoleuca*– and hybrid-type individuals of these populations show higher proportions of topologies congruent with color (*O. melanoleuca*) introgression, mostly in the ASIP region (Pontic, 11.8% more topologies congruent with introgression in *O. melanoleuca* and hybrid phenotypes than in parental phenotypes; Caspian, 6.1%), much less across chromosome 20 (Pontic, 1.7%; Caspian, −0.3%), and not at all genome wide (fig. S20). This suggests that the plumage color polymorphism in these populations is a result of genomically very localized and likely old introgression of white coloration, similar to introgressed wing-pattern variation in *Heliconius* butterflies (*4*, *46*).

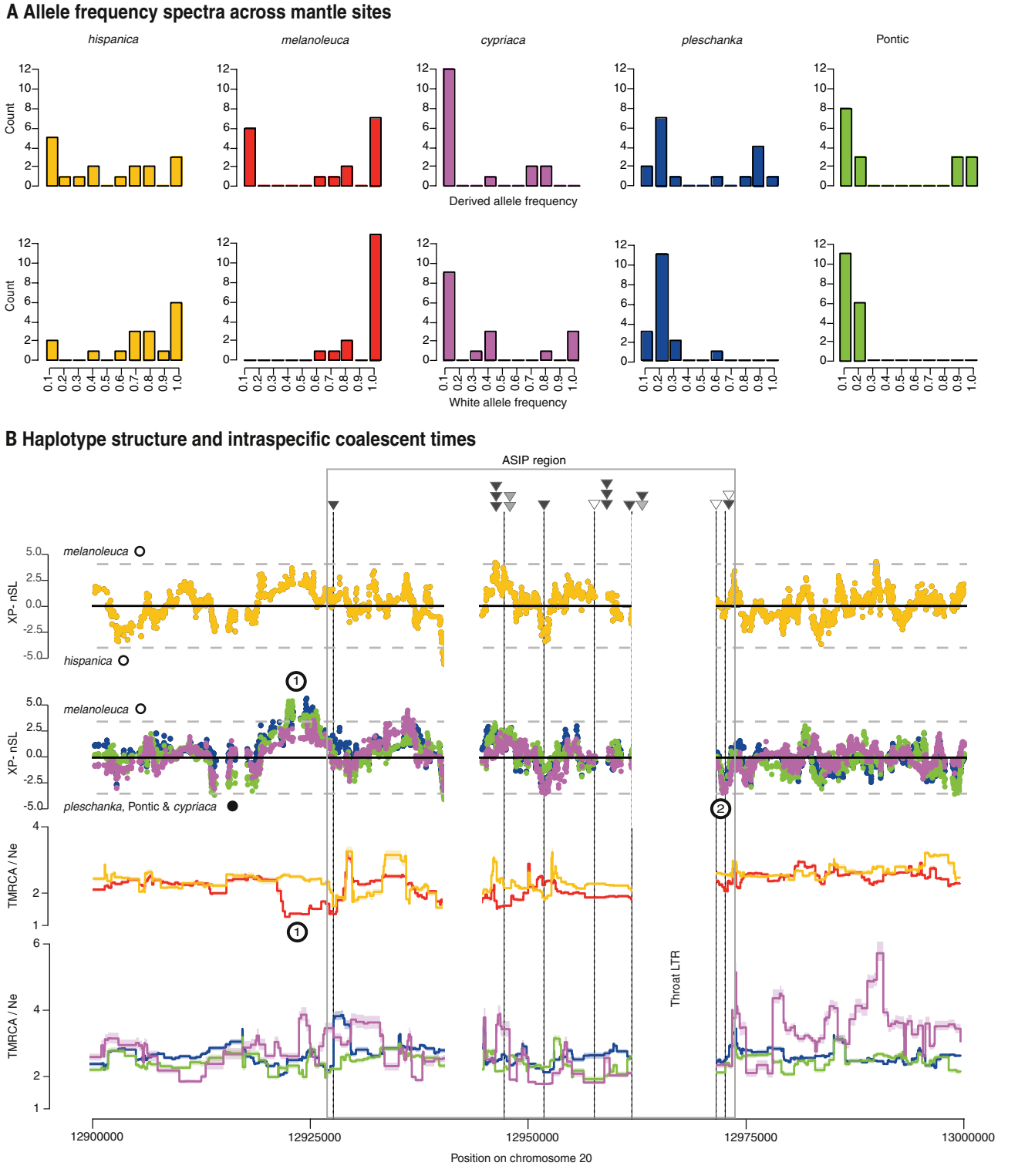


Fig. 6. Evidence for positive selection on ASIP variation. (**A**) Unfolded site frequency spectra for the 17 mantle loci polarized according to ancestry (top) or mantle coloration (bottom). Stronger shifts toward species' colorations than toward ancestral or derived variants (notably strong in *O. melanoleuca*, absent in *O. pleschanka* and the Pontic lineage) are in line with positive selection on coloration. (**B**) (Top) Cross-population haplotype statistics [XP-nSL; (*94*)] between *O. melanoleuca* and *O. hispanica* and between *O. melanoleuca* and all black-mantled species (mantle coloration is indicated by the color of circles besides the species labels). Positive and negative peaks indicate selection in the lineage denoted to the left of the panels. Colors indicate the species to which *O. melanoleuca* was compared (colors follow those in Fig. 1A). (Bottom) ARG-based TMRCAs (*37*,*38*) within species scaled by effective population sizes (estimated based on nucleotide diversity), with low values indicative of positive selection. Footprints indicating positive selection in *O. melanoleuca* (labels 1) and (weaker) in the region flanked by two derived black variants in *O. pleschanka* and the Pontic lineage (label 2) are highlighted. Dashed lines show the positions of mantle loci, with triangles at the top indicating sites that, in the complex's ancestor, were black (black), polymorphic (gray), or white (white). Lines depict multiple mantle loci where the latter are in close proximity.

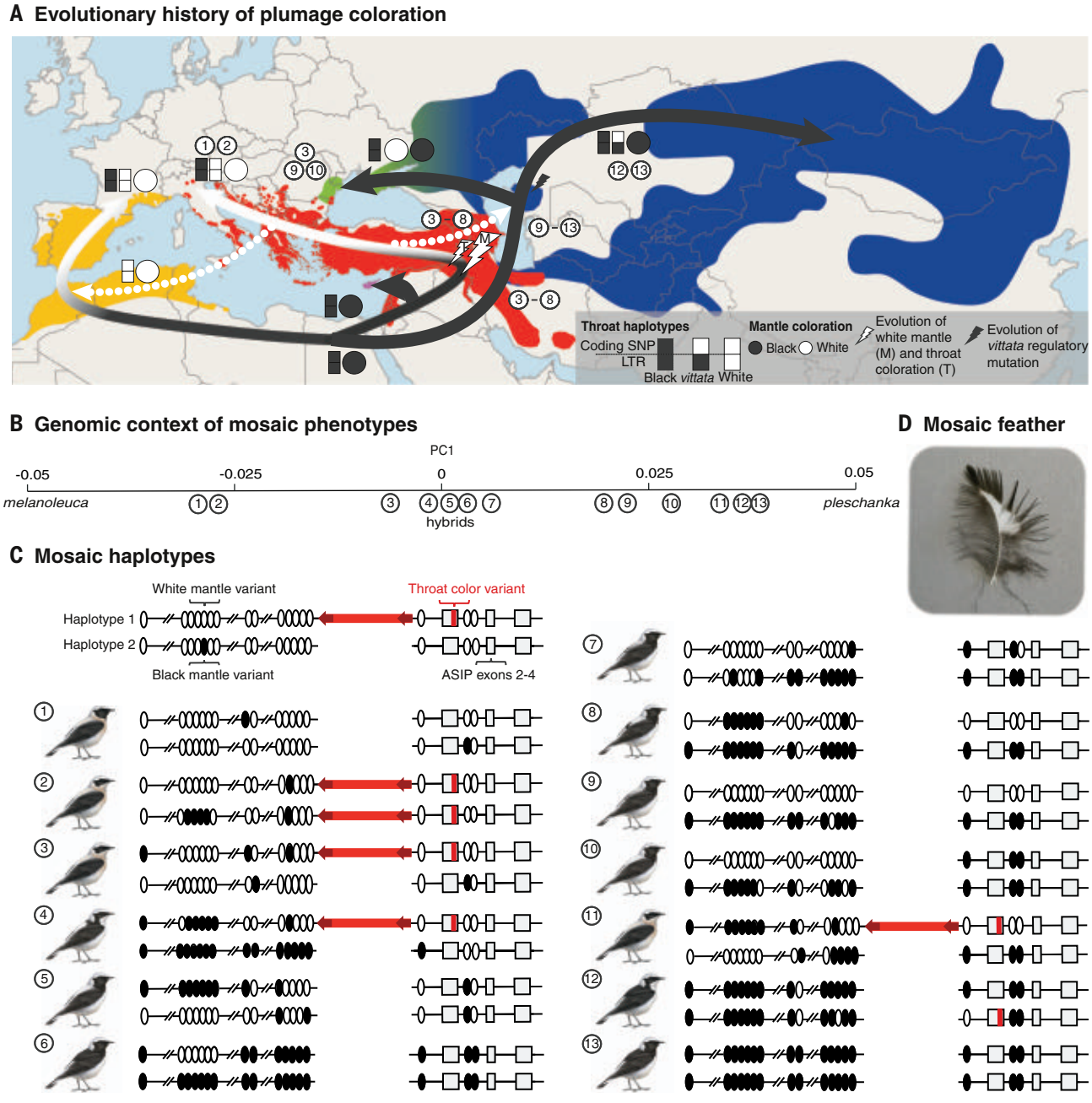


Fig. 7. Evolutionary history of wheatear coloration and recombined phenotypes and haplotypes. (**A**) The evolutionary histories of throat and mantle coloration are shown by arrows in the distribution maps (colors follow those in Fig. 1A). Black arrows depict ancestrally black mantle and throat coloration in *O. pleschanka*, the Pontic lineage, and *O. cypriaca*, with mantle coloration fixed in all three lineages. White mantle and throat coloration evolved in the *O. melanoleuca* ancestor after the split from *O. cypriaca* (white lightning arrows, which transition to white arrow in *O. melanoleuca* range); white mantle coloration is fixed in this species, whereas throat coloration is polymorphic. Then white coloration introgressed from *O. melanoleuca* into *O. hispanica* (dotted arrow), leading to fixed white mantle and polymorphic throat coloration. Further east, the throat SNP introgressed from *O. melanoleuca* into *O. pleschanka* (dotted arrow) and, likely together with a new regulatory mutation (black lightning arrow), led to the evolution of the *vittata* phenotype. Locations where phenotypes depicted in (C) can be observed are indicated by circled numbers. Arrow lengths are not at scale with evolutionary timings. Although the evolutionary processes are projected on current species distributions, geographic locations are not to be interpreted. (**B**) Phenotypes' typical ancestries (several possible) are shown on PC1 for phenotypes and haplotypes from (C). (**C**) Selected range of phenotypes together with examples of the underlying ASIP haplotypes. Ovals depict mantle loci, throat variation is shown in red, and ASIP exons are shown as squares. The depicted haplotypes illustrate the fine-scale recombination of ASIP variation. At mantle loci, variation blends into a continuum of almost exclusively white (number 1), over different degrees of alternating black and white (numbers 2 to 12), to exclusively black variation (number 13). White-throated individuals exhibit at least one haplotype containing the throat LTR (red arrows) and white throat SNP variant (red box), except *vittata* (number 12). (**D**) Mosaic black-and-white back feather, as observed in various hybrid-type phenotypes.

Color mosaics can be observed down to the level of single feathers. In intermediate phenotypes, single mantle feathers commonly display a mosaic of black-and-white coloration (Fig. 7D). This suggests that mosaic ASIP haplotypes may result in unstable ASIP expression.

New mutations underpin genus-level convergence

The plumage coloration in the *hispanica* complex evolved convergently also at the genus level and beyond (*22*, *23*). Therefore, our final interest was to understand whether the same variants as within the *hispanica* complex underpin convergent evolution across all wheatears and related species. To this end, we interrogated alignments of whole-genome resequencing data of 50 individuals from 44 species of wheatears and their close relatives (*22*). At all sites associated with throat and mantle coloration (including the throat LTR) for which data was available, we only observed the ancestral variants in species outside the *hispanica* complex (fig. S21). This highlights that the variation underpinning plumage coloration at the genus level differs from the variation associated with these colorations in the *hispanica* complex. Thus, consistent with results from other bird species (*2*, *12*, *15*), whereas the shorter-term color evolution in the *hispanica* complex is underpinned by interspecific exchange of genetic variants, the longer-term convergent evolution of plumage color diversity beyond the *hispanica* complex is due to new mutations (*19*).

Evolution of recombined phenotypes within the constraints of a monogenic architecture

Wheatear coloration shows how phenotypes can be decoupled between body parts despite being regulated by a single gene. The genetic architecture involving modular regulation of a single gene sets wheatear coloration apart from what is known from other bird systems, where recombined color phenotypes are underpinned by multiple genes (e.g., *2*, *12*, *47*) or a single gene produces uniform coloration across multiple feather tracts (e.g., *11*, *48*, *49*). In the *hispanica* complex, independent colorations of multiple body parts are determined by distinct regulatory modules of ASIP, similar to that in mammals (*17*, *34*, *50*). Also, the grouping of mantle loci in blocks (Fig. 2B) and the alternative transcription start sites observed among isoforms (fig. S12) reflect a pattern known from mammals, where this pattern represents alternative untranslated first ASIP exons (*17*, *50*). The involvement of a TE in ASIP regulation was hitherto only hypothesized for bird coloration (*51*) but has been documented for mammalian ASIP regulatory modules (*36*, *51*). The modular regulation of wheatear coloration is thus similar to what is documented from mammals and insects (*9*, *17*, *50*, *52*) and reflects a likely universal modular regulatory architecture of ASIP that enables decoupled melanin expression between body parts.

However, modular gene regulation alone may not always grant independent evolution of the associated phenotypic elements. The supposed evolutionary history of the *vittata* phenotype suggests that the interplay of modules' linkage with selection against introgression may have prevented opportunities for the recombination required to decouple the evolutionary histories of different body parts. We deem it likely that, in *O. pleschanka*, selection against white mantle coloration prohibited the introgression of the throat LTR that is usually linked to white mantle variation. Therefore, selection limited the opportunity for recombination of the throat LTR onto haplotypes carrying black mantle variation in *O. pleschanka*. Consequently, the evolution of the *vittata* phenotype appears to have required a new element up-regulating ASIP expression in throat plumage. Hence, with selection limiting introgression and opportunities for recombination, the modularity of ASIP expression was not sufficient to completely resolve the pleiotropy between throat and mantle coloration.

For preexisting variation to contribute to phenotypic diversity underpinned by a single gene, sufficient opportunity for recombination to decouple regulatory modules turns out to be key. Together with previous insights from dogs, ladybird beetles, and *Heliconius* butterflies (*9*, *17*, *18*), the evidence from wheatears establishes that, in the long run, even within the constraints of monogenic architectures, fine-scale recombination can integrate linked regulatory modules from divergent origins into new recombined haplotypes. This appears to imperatively involve long-term recombination, be it within species (*9*, *17*, *18*) or their ancestors (*17*) or in hybrid zones (wheatears). In this context, the almost seamless integration of black and white back variation upstream of ASIP in hybrid wheatears (Fig. 7C) demonstrates the power of combinatorial processes in the evolution of phenotypic diversity and underscores the importance of hybrid zones as melting pots that enable evolutionary tinkering.

Convergent plumage color evolution involved diverse types and origins of genetic variation

The patterns, processes, and genetic material involved in wheatear color evolution paint an intricate evolutionary history, exemplifying that "either-or"–type questions that shaped evolutionary genomics for over a decade (*53*–*56*) may rarely find single, entirely black or white answers. Wheatear color evolution involves a single gene, yet different regions associated with this gene; it involves preexisting and new variation, protein coding, and regulatory variation (*53*–*55*), and it is based on both point mutations and structural variation (*51*, *56*). This highlights that the diversity of wheatear coloration is underpinned by a mosaic of genetic variation, evolutionary origins, and evolutionary histories.

We emphasize that a study system spanning variable evolutionary timescales and hybridization contexts was key to obtain insights into evolution from preexisting variation beyond the latter's contributions to phenotypic diversity. Our findings suggest that the role of preexisting variation in wheatear plumage color evolution was limited by its availability and the constraints imposed by a monogenic architecture. First, extant hybrid zones showcase how long-term recombination can break up even tightly linked variation at the fine scale required to reshuffle modular variation into recombined haplotypes. Yet, the resulting phenotypes do not contribute to plumage color diversity outside hybrid zones, such as it was achieved by *O. melanoleuca*–*O. hispanica* hybridization about 400 kyr ago. Second, past hybridization with *O. melanoleuca* had distinct outcomes in *O. hispanica* and *O. pleschanka*. It shaped all color variation of extant *O. hispanica* but contributed only one component for the convergent evolution of the white-throated *vittata* phenotype in *O. pleschanka*. At longer timescales at which preexisting variation was no longer available—that is, across the genus—plumage color evolution was instead driven by new mutations. Similarly, new mutations seem to have been involved where contributions of preexisting variation were likely constrained by linked selection, as seen in *vittata*. Together, these results offer a nuanced perspective on how genetic architectures constrain phenotypic evolution from preexisting variation and, more broadly, how genetic architectures can shape evolutionary trajectories over time.

The detailed insights into the roles and interactions of genetic architecture and diversification processes in phenotypic evolution gained from wheatear coloration may be relevant to our understanding of the reshuffling of preexisting and new genetic variation during the diversification of a wide range of organisms. Our results show that evolution recruits variation from all sources it has at hand, genetic architecture permitting. At deeper timescales, mutations were both frequent enough and required to give rise to convergent plumage phenotypes. Future comparisons of evolutionary dynamics across a wide range of phenotypes and organisms will be required to understand whether this is only possible due to extraordinarily dynamic phenotypes and genetic architectures, or, indeed, mutations may be less limiting than assumed to date (*57*), and/or the regulatory effects of new TE insertions play an even more important role in phenotypic evolution than already commonly acknowledged (*51*, *58*). For example, we suspect that, in line with TEs' effects on color evolution across the

animal kingdom (*51*), extended studies of structural variation across the genus will reveal an important role for TEs in the genetic architecture of coloration in other wheatears and their relatives. Still, preexisting variation and permeable species barriers in wheatears, as in many organisms (*59*), prove to provide abundantly used substrate for adaptation and phenotypic evolution. This highlights humankind's responsibility to conserve a maximum number of species and genetic variation to equip biodiversity with the molecular prerequisites to adapt to the increasing challenges encountered in the Anthropocene.

Materials and methods

Whole-genome resequencing

To obtain a representative sampling of genetic and phenotypic variation across the *hispanica* complex, we collected samples from the major parts of the species' distribution (Fig. 1C). To optimize the power of admixture mapping, hybrid zones were densely sampled. During fieldwork dedicated to the present project, we collected whole blood into >96% ethanol, Queen's lysis buffer, or DNA/RNA Shield (Zymo). Samples obtained from museum collections were either blood or muscle tissue in ethanol, or toe pads collected from skins (table S1).

We extracted DNA from blood and tissues using the DNeasy Blood and Tissue Kit (Qiagen) or the MagAttract HMW DNA kit (Qiagen) following the manufacturer's protocols, with the exception of an adapted digestion of blood samples (*60*). We extracted DNA from toepads using the QIAmp DNA Micro Kit (Qiagen) with a digestion protocol adapted to ensure sufficient DNA quantities (https://www.protocols.io/view/dna-extraction-protocol-for-historical-toe-pad-sam-dm6gpwrdplzp/v1). For quality control, DNA concentrations were determined with a Qubit fluorometer (dsDNA BR assay, Thermo Fisher Scientific), and DNA integrity for linked-read sequencing was evaluated on a TapeStation with the Genomic DNA Screen Tape (Agilent) or on a FEMTO Pulse CE instrument (Agilent). To obtain high quality gametic phase information for a reference haplotype panel, we sequenced 238 individuals using linked-read technology (*60*). Linked-read sequencing libraries were prepared using Chromium Genome library kits (10X Genomics) (table S1). Short-read sequencing libraries for 139 additional individuals for which fresh material was available were prepared with either the Illumina DNA Prep Kit, the Illumina DNA PCR-free kit, or following Baym *et al.* (*61*); sequencing libraries for 12 individuals with degraded DNA were prepared with the xGen ssDNA & Low-Input DNA Library Prep Kit (Integrated DNA Technologies) (table S1). Apart from the seven libraries reported in Lutgen *et al.* (*60*) that were sequenced on an Illumina HiSeq-X instrument, all libraries were sequenced on an Illumina NovaSeq6000 instrument with a target coverage of ca. 15x. After the removal of individuals with a coverage below 8x, this resulted in a dataset comprising a total of 379 individuals (tables S1, S13).

Data preparation

To prepare whole-genome resequencing data for analysis, linked-read data and standard short read data were first treated separately, following the GATK 4.1.4.1 (*62*) best practices pipeline where appropriate. We aligned linked reads to the *O. melanoleuca* reference genome (*63*) using phase-aware alignment in Lariat as implemented in LongRanger 2.2.2 align (*64*). To obtain a SNP exclude set for base quality score recalibration (BQSR), SNPs were then called and genotyped in GATK with HaplotypeCaller as depicted in fig. S22. To include only high-confidence SNPs in the SNP exclude set for BQSR, we retained only SNPs with mapping quality > 40, Fisher strand phred-scaled p-value >60, SNP quality score >20, mapping quality rank sum value > -12.5, read pos rank-sum test value > -8.0 and quality by depth > 5. We retained only biallelic SNPs with at least one homozygous reference and one homozygous alternative genotype or with a minimum of three observations of reference or alternative alleles. BQSR was performed in GATK followed by final SNP calling with HaplotypeCaller. Short reads were adapter-trimmed in fastp 0.20.0 (*65*) before mapping them to the *O. melanoleuca* reference genome (*63*) using bwa-mem 0.7.17 (*66*). We then marked duplications with PicardTools 2.9.1 (http://broadinstitute.github.io/picard) and directly performed BQSR with the SNP exclude set obtained from the geographically and taxonomically representative linked-read data (n=235, table S1). SNPs were called with HaplotypeCaller. To obtain a joint VCF file with unphased SNP data for linked-read and short-read data, we imported the resulting gVCF files into a single genomic database and genotyped SNPs for both data types. The resulting VCF constituted the basis for analyses not requiring gametic phase, such as GWAS.

To infer gametic phase for phase-aware analyses, we leveraged the linked-read information from 235 individuals to establish a high-quality reference haplotype panel (*60*) with which we then supported the statistical phasing of short-read data. To this end, for each individual we phased heterozygous genotypes obtained in the last step either with (i) linked-read-barcode data using HapCUT2 (*67*) for individuals with linked reads available or (ii) paired-end information using whatshap v.1.1 (*68*) for the remaining individuals. To include all originally called SNPs in the individual VCF files, we parsed the phased and unphased data for each individual to reintroduce SNPs homozygous in the individuals. We then parsed individual vcf files into either of two population vcf files, separated by linked-read and short-read data. To establish the final reference haplotype panel, for linked-read data we retained only high-quality SNPs that had a minimum sequencing depth of 5 and minimum genotype quality of 20 in at least 75% of individuals. This left a total of 69,855,099 SNPs (table S13). We then established the gametic phase of phase sets statistically using Shapeit 4.2 (*69*), assuming that phase sets established beforehand based on linked-read info are correct (–use-PS 0). Finally, we introduced the resulting haplotypes as reference haplotype panel into the statistical phasing of short-read data in Shapeit (–use-PS 0.0001). To avoid biases introduced by population stratification we performed this statistical phasing separately for each of the 11 populations (one *cypriaca*, four *melanoleuca*, two *pleschanka*, two Pontic, and two Caucasus populations) using a specific reference panel consisting of individuals from populations closest related as determined from the PCA (Fig. 1D). Short-read data were not filtered prior to statistical phasing, as with the reference haplotype panel and an average sequencing coverage of 18 × per individual, imputation of missing data are expected to yield adequate genotypes.

For analyses that require an outgroup, we added unphased data from three *O. deserti* individuals to the phased dataset. For each of the three *O. deserti*, we generated allsites vcf files in GATK and extracted genotypes for the sites contained in the phased dataset using GATK's SelectVariants. Using bcftools we then merged the outgroup data with the phase dataset and removed *O. deserti*-specific indels that contained no data in the phased dataset. To reconstruct the ancestral states of the *hispanica* complex and the *O. deserti* outgroup, additional allsites vcf files were generated for two *O. monacha* and three *O. oenanthe* serving as distant outgroups.

A characterization of all data sets used for analysis is provided in table S13. Individuals used for specific purposes or analyses are listed in table S14.

Reconstruction of the *hispanica* complex's evolutionary history

We first aimed to obtain insights into the genetic structure of the *hispanica* complex without prior assignment of individuals to populations. To this end, we performed a PCA on genotype frequencies of all 379 individuals using smartpca (*70*) and inferred co-ancestries among individuals and based on this individual clustering in fineSTRUCTURE (*28*).

For PCA, we used the SNPs contained in the genome-wide phased data set. To exclude data affected by poor mapping in repeat regions, selection on genes, and linkage disequilibrium, we removed (i) repeats based on the annotation using the repeat library from Peona *et al.* (*63*),

(ii) gene regions using the annotation bed file published in Peona *et al.* (*63*), and (iii) regions with high linkage disequilibrium (–indep-pairwise 50 5 0.2) using plink v.1.9 (*71*). To ensure that the PCA is not biased by our sampling's high load with individuals from hybrid zones, in addition to the standard PCA including all individuals (fig. S23), we ran smartpca first with 12-15 individuals per each of the five taxa (*hispanica:* 14; *pleschanka:* 15; Pontic: 15; *melanoleuca:* 15; *cypriaca:* 12), with individuals chosen from the core clusters of the former PCA (hereafter referred to as "reference individuals;" fig. S24) and projected the genotypes of the remaining individuals onto the inferred axes using smartpca. Two individuals were strong outliers, with each producing an axis separating it from all other individuals (not shown). We therefore removed these individuals from the final analyses. To estimate the level of differentiation between lineages, we computed pairwise differentiation based on Weir and Cockerham's F_{ST} implemented in vcftools (*72*).

To investigate how ancestries transition across the hybrid zones (Iran, Caucasus, Black Sea, and Caspian), we performed one-dimensional sigmoid geographic cline analyses in the HZAR package in R (*73*). To this end, first, we for each hybrid zone first identified the principal component axis (PC) that best separates parental lineages. The respective PC axes were then used as a proxy for hybrid index by rescaling PC scores to between 0 and 1. Second, we estimated Bayesian hybrid indices in bgchm (*74*) To this end, we for each lineage pair extracted ancestry-informative markers (AIM), with AIM defined as SNP with F_{ST} as estimated in vcftools (*72*) exceeding 0.7. For comparisons exceeding 10,000 AIM, we randomly sampled 10,000. This resulted in each 10,000 AIM for *melanoleuca-pleschanka* across Iran, *melanoleuca-pleschanka* across the Caucasus, and *melanoleuca*-Pontic (Black Sea), and 4,369 AIM for the Pontic-*pleschanka* (Caspian) comparison. Based on these AIM we then estimated Bayesian hybrid indices using the est_hi function in bgchm (*74*), using reference individuals as parentals (table S14, fig. S24). For each hybrid zone, we then projected samples along a one-dimension transect using the distm function implemented in the geosphere R package (https://github.com/rspatial/geosphere). Specifically, we calculated the great circle distance between each individual and the westernmost parental individual averaging across samples for the latitude. We assigned samples to populations along the transect by grouping individuals within a 100 km radius into a single population. Next, we fit three geographic cline models with HZAR using hybrid indices per sampling location as the dependent variable and the geographic distance along the one-dimensional transect (in 100 km) as predictor. We fitted clines using the default MCMC chain length of 100,000 steps with 10,000 burn-in steps and selected the best model by comparing the support for then null model and the three cline models using AIC. We plotted the best model having an AIC threshold score >2.

To investigate admixture and infer population clusters based on individuals' co-ancestries we leveraged our phased data to trace excess haplotype sharing using the powerful approach implemented in fineSTRUCTURE v.4.1.1 (*28*). We visualized the results of this analysis as a clustering tree, haplotype frequency-based PCA, and co-ancestry matrix. For fineSTRUCTURE analyses we corrected excess relatedness among individuals by removing each one individual from nine individual pairs with above second degree relatedness (>0.2; vcftools -relatedness2), because their inclusion would bias population clustering given the extremely high excess of recent haplotype sharing, and we thinned the phased whole-genome dataset using vcftools (*72*) to retain 1 SNP every 500 base pairs (-thin 0.5). Thinning was performed to reach achievable computing times, with parameters based on a comparison of fineSTRUCTURE runs using complete and thinned data sets from (micro)chromosomes 15-28. FineSTRUCTURE was run with a recombination map estimated from *melanoleuca* data using pyrho v.0.1.7 (*75*). Because population structure can confound LD-based recombination rate estimates, we restricted recombination rate estimation to a single population of *melanoleuca* from Greece (N=13). Population size trajectories were estimated with MSMC2 for five individuals from this population to account for demography within the pyrho inference. Optimal hyperparameters were determined through 50 rounds of simulations for all possible combinations of parameters from block penalties of 10, 15, 20, 25, 30, 35, 40, 45, and 50 and window sizes of 30, 40, 50, 60, 70, 80, and 90. A per-generation mutation rate of $4.6 \cdot 10^{-9}$ from collared flycatcher (*76*) was applied to convert population-scaled recombination rate to cM/Mb. The pyrho output was converted to genetic map format with a custom script, and the genome-wide average recombination rate was calculated by taking the average rate between SNP pairs, weighted by the physical distance between SNPs.

To reconstruct the *hispanica* complex's speciation history, we inferred the species tree and estimated demographic trajectories and split times based on the multiple sequentially Markovian coalescent (MSMC) and based on the ancestral recombination graph (ARG). We estimated the species tree based on non-overlapping 10 kbp windows in the reference individuals. Data from three *O. deserti* were included to root the species tree. We only considered windows meeting the following criteria, following Alaei Kakhki *et al.* (*22*): minimum read depth of 5, minimum 50% of the window covered by data, and maximum 20% missing data per site (n=79,127). To include only windows without recombination but free inter-locus recombination, we excluded all windows with recombination signals ($P \leq 0.05$) as inferred from the pairwise homoplasy index Phi (Φw) estimated in PhiPack 1.1 (*77*) and enforced a minimum distance of 10 kbp between consecutive loci, as at this distance no linkage disequilibrium occurs in flycatchers (*78*) (N=29,792). Based on this data, we estimated the species tree with multiple individuals forced together in the species tree (-a option; Fig. 1E), or keeping all individuals as leaves (without the -a option; fig. S3) using the multispecies coalescent in ASTRAL-III (*79*) based on maximum likelihood gene trees estimated in IQ-TREE 2.1.2 (*80*) using a GTR+I+G substitution model and one thousand ultrafast bootstrap approximations. We estimated the MSMC using MSMC2 (*81*) using data of five individuals of each of the five taxa. Where possible, we included individuals with linked-read data available to ensure the best phase information. To make use of the phase information, we first generated sample-specific VCF and mask files using the bamCaller.py from msmc-tools, and into the resulting VCF files injected phase information from our phased VCF file. SNPs without phase information were excluded. MSMC2 was run to evaluate coalescence between haplotypes within individuals for each population (-I) with default time intervals (-p $1 \times 2 + 25 \times 1 + 1 \times 2 + 1 \times 3$) and a ratio of recombination over mutation rate (-r) of 5.65 based on the flycatcher mutation rate of $4.6 \cdot 10^{-9}$ per bp per generation (*76*) and a mean autosomal recombination rate of $2.6 \cdot 10^{-8}$ (as estimated for *melanoleuca*, see above). To convert generations into years, we assumed a generation time of two years.

To infer the relative timing of lineage splits based on coalescence times, we leveraged the power of phased data and ARGs. ARGs were inferred from phased data using tsinfer v. 0.3.1 (*37*) for the reference individuals. We ran tsinfer with a mismatch ratio of 0.1 and provided a uniform recombination rate for each chromosome, using the mean recombination rate estimate for *melanoleuca* (table S15) with pyrho (see above). In this way, we accounted for broadscale interchromosomal variation in recombination rate, where the average recombination rate increases with decreasing chromosome size. We then estimated the age of ancestral nodes in the inferred trees with tsdate v. 0.1.7 (*38*), using the variational gamma method of dating. For each chromosome we provided a mutation rate of $4.6 \cdot 10^{-9}$ per bp per generation (*76*), while a chromosome-specific N_e was estimated from chromosome-level genetic diversity. We then extracted the TMRCA for each pair of haplotypes across the full set of trees based on the ages estimated by tsdate. To obtain genome-wide estimates of divergence times, we calculated the logarithmic mean TMRCA, weighted by tree span, for haplotype pairs of each interspecies comparison. Mean TMRCAs were estimated for all trees genome-wide, restricted to trees

overlapping with genic regions, and for all trees on chromosome 20. We then obtained 95% confidence intervals around the mean values of TMRCA by performing 100 bootstrap replicates, in which we resampled haplotype pairs with replacement and re-estimated the mean TMRCA.

To reconstruct the *hispanica* complex's introgression history, we investigated genome-wide D statistics and f4-ratios using Dsuite (*82*). For these analyses we specified *O. deserti* as outgroup (*22*).

Phenotype scoring

To obtain phenotype scores of throat, mantle (here referring to mantle plus back) and neck coloration for GWAS, we categorized plumage traits based off pictures of field-caught or museum specimens. We scored throat coloration as discrete black or white. For mantle and neck side coloration, intermediate phenotypes are frequent. Hence, we divided these plumage traits into four ordinal categories: white (1), white with some black feathers (2), black with some white feathers (3), and black (4). To visualize the transition of mantle and neck side coloration across geography and admixture proportions, we estimated a phenotypic hybrid index based on the average between individuals' mantle and neck side scores (Fig. 1H, fig. S7).

GWASs

To identify genetic variants associated with polymorphic throat, and divergent mantle and neck side coloration, we conducted GWAS for each of the three plumage color traits. To avoid eliminating relevant SNPs, we preferred a strategy of initially restricting filtering to a minimum but checking for adequate data quality of significant sites recovered by GWAS post hoc. To this end, we determined filtering thresholds based on parameter distributions of Fisher Strand (FS), Strand Odds Ratio (SOR), RMS Mapping Quality (MQ), Mapping Quality Rank Sum Test (MQRankSum), Quality by Depth (QD), Read Position Rank Sum Test (ReadPosRankSum), and combined depth across samples (INFO/DP) and filtered for FS>40.0, SOR>3, MQ<4, MQRankSum>5, QD<2.0, ReadPosRankSum=4, and INFO/DP>13000 using bcftools v1.10 (*72*). We further removed (i) individuals with missing phenotypic information, (ii) females (because they hardly express the phenotype and therefore scoring is difficult or impossible), (iii) individuals with a sequencing coverage below 8x, and (iv) individuals that were outliers in the PCA (ones that formed a principal component by themselves). The final GWAS data set comprised 336 (58 white and 278 black) individuals for throat coloration and 230 individuals for mantle and neck side coloration. Because mantle and neck coloration are fixed in the parental species and only recombine in hybrids, we only included (hybrid) populations that comprise phenotypically admixed individuals in the respective GWAS (Fig. 1H and fig. S7). We conducted GWAS using linear mixed models to account for relatedness as implemented in GEMMA v0.98.5 (*30*). We ran all models with default settings, removing (i) non-polymorphic SNPs, (ii) SNPs with a missingness above 5%, (iii) SNPs with a minor allele frequency below 1%, and (iii) SNPs with an r^2 above 0.9999, which left 40,180,174 SNPs (of originally 162,197,710) SNPs. Throat color variation was almost entirely explained by two SNPs in consecutive codons in ASIP exon 2. Genotypes between the two SNPs were perfectly correlated except for genotypes of the second SNP that showed a third, non-functional indel variant whose pseudogenizing effect fit the associated phenotypes (n=8).

We followed up all GWAS on the global datasets described above by running GWAS split by species and hybrid zone, respectively. For throat coloration, we conducted GWAS using linear mixed models accounting for ancestry in *melanoleuca* (n=46, n_{white}=21, n_{black}=25) and the Iranian hybrids (n=87, n_{white}=17, n_{black}=70). We refrained from modeling GWAS for *hispanica* and *pleschanka* given the limited sampling sizes in *hispanica* (n=11, n_{white}=5, n_{black}=6) and of the white-throated *vittata* phenotype in *pleschanka* (n=2). With these very limited sample sizes, GWAS in GEMMA that model traits as additive did not have the power to identify SNPs associated with the dominant throat color architecture. We therefore reran GWAS under a dominant genetic architecture in angsd 0.940 (*83*). For mantle and neck-side coloration, we performed GWAS on four sets of admixed individuals: Iran (n=87), Black Sea (n=65), Black Sea and Caspian (n=93), and Caucasus (n=50).

Identification, genotyping, and phasing of structural variants

Because ASIP usually affects color phenotypes through regulatory changes that can involve structural variants (SVs) (*17*), we updated the repeat annotation of chromosome 20 with a manually curated version (*84*) of the *O. melanoleuca* library (*63*) (data S3) and investigated the ASIP region for SVs associated with throat coloration. To identify SVs on chromosome 20, we ran Smoove 0.2.8 (https://github.com/brentp/smoove) using all the short-read libraries available. The genotypes of 291 individuals were directly inferred by Smoove (the remaining 44 individuals could not be genotyped due to insufficient sequencing coverage). SVs were then manually curated using Samplot 1.3.0 (*85*) and Plotcritic 1.01 (*86*) to remove false positives. We overlapped the SV annotation with the repeat annotation performed with RepeatMasker/4.1.0 using the new version of the *O. melanoleuca* library. An SV linked to the throat SNP ('throat LTR') was found to be a full-length endogenous retrovirus (LTR retrotransposon) with a target site duplication of 5 bp (ERVL oenMel1-22) (fig. S10). To investigate the presence of the throat LTR in the white-throated *vittata* phenotype of *pleschanka* based on a genome assembly, we investigated the insertion site in a published draft genome assembly of this phenotype [(*60*); ENA biosample SAMEA6823426]. To phase the throat LTR and throat SNP, that is, to infer whether in heterozygous individuals the LTR insertion locates on the same haplotype as the white throat SNP variant, we interrogated linked-read barcodes in 38 heterozygous individuals with linked reads available in the ASIP region. To this end, we first selected all reads supporting the two throat SNP variants and then collected the reads that shared those barcodes separately for the two variants. We then mapped the reads supporting the black and white variants separately to the genomic region comprising the ASIP gene and the throat LTR and in both instances counted the reads mapping to the LTR.

Statistical modeling of genetic architecture

We aimed to understand how variants inferred by GWAS interact to form the associated phenotypes. To this end, we modeled mantle and neck side variation with an additive effect by, for each individual counting the number of white alleles (alleles associated with predominantly white phenotypes) across mantle loci. To model the phenotypic effect of white as dominant, we counted the number of mantle loci with at least one white allele for each individual, assuming that the phenotypic effect of one single white allele is the same as for two white alleles for this site. To model black as dominant, we did the same analysis, but counting the number of sites with at least one black allele for each individual. We then used these genetic effects to predict mantle or neck side phenotypes using Bayesian linear regression models as implemented in the brms package version 2.20.4 (*87*) in R version 4.3.3. We fitted Bayesian linear regression models with the family categorical, using four chains each running 4,000 iterations with a warm-up of 2,000 iterations. We then used the loo_compare() function implemented in loo package to identify the best performing model using leave-one-out (LOO) cross-validation. The cross validation implemented in the loo_compare() function performs pairwise comparisons of each model to the model with the highest estimated log predictive density (ELDP) and generates the difference in ELDP and the associated standard error (*88*).

As evidenced in Fig. 2C the throat LTR and SNP genotypes explained all color variation observed in *hispanica* and *melanoleuca*. We consequently refrained from further statistical modeling (Bayesian models require variation).

Gene expression analysis

To confirm the regulatory nature of GWAS hits and corroborate the genetic architecture of throat and mantle plumage coloration, we investigated ASIP expression and transcript diversity in wheatears of alternative phenotypes. As representatives of alternative plumage coloration, we targeted white-mantled but alternatively white- or black-throated *melanoleuca* and black-mantled and black-throated *cypriaca*. Two white-throated *melanoleuca* individuals were caught in Greece, two black-throated *melanoleuca* individuals were caught each one in Greece and in Cyprus, and three *cypriaca* individuals were caught in Cyprus. Birds were caught between August 22nd and September 18th, 2024, during (*cypriaca*) or after (*melanoleuca*) post juvenile or post-breeding moult when the birds are or have been in a physiological state conducive of feather regrowth and first year birds already moulted feathers to adult-type coloration. Immediately after capture, the birds were transported to bird care facilities at Pafos Zoo (Paphos, Cyprus) or Attica Zoological Park (Athens, Greece).

To characterize ASIP transcript variation and expression during feather growth—that is, when pigments are deposited in the feathers—we induced the regrowth of feathers. To this end, we pulled feathers from the two feather tracts of interest, throat and mantle, as well as from the shoulder and the belly, where both species express black (shoulder) and white colored plumage (belly). Shoulder and belly feathers served as a control to differentiate whether gene expression indeed differed between feathers of different coloration and not between species. Feather regrowth proceeded heterogeneously between species, individuals, and feather tracts in regard to both the onset of regrowth and whether or not feathers regrew. Given that throat and especially mantle feathers do not exhibit uniform color when fresh but particularly in *melanoleuca* show banded color patterns (fig. S25), we pulled feathers at sizes during which we expected them to express the respective white or black color for each feather tract (based on the patterns of the initially pulled feathers; fig. S25). Pulled feathers were immediately stored in liquid nitrogen until further storage and transportation and final storage at -80°C.

Next, to investigate associations of gene expression in regrown feathers with black and white coloration in the different feather tracts, we performed single-molecule real-time isoform sequencing (Iso-Seq) of total RNA on Pacific Biosciences' Revio long-read sequencing platform. To this end, from the available regrown feathers (table S16) we made selections of each 3-9 regrowing feathers to go into ten RNA pools representing alternative colorations and feather tracts (table S17). For *melanoleuca* mantle feathers, we generated three pools with feathers of different sizes and visible coloration (table S17) to ensure that we capture the stage at which white is expressed (fig. S25). From these feather pools we extracted total RNA using the RNeasy Universal kit (Qiagen) following the manufacturer's instructions. The tissues for RNA extraction consisted of the feather tips embedded in the skin that include the follicular sheet, the epidermal collar, part of the dermal papilla, and the pulp of the feather. RNA extracts were cleaned up with the RNeasy mini kit (Qiagen). RNA quantities and qualities were controlled on a Qubit fluorometer (Thermo Fisher Scientific) or an RNA ScreenTape on a TapeStation (Agilent). To prepare the RNAs for sequencing they were transcribed to cDNA and barcoded individually using the Iso-Seq express 2.0 kit (Pacific Biosciences). To maximize sequencing yield, we constructed three sequencing libraries using the KINNEX kit (Pacific Biosciences), with pools grouped into KINNEX libraries as described in table S17. The three resulting libraries were then sequenced each separately on one SMRT cell 25M (according to table S17) at the NGS platform of the University of Berne, Switzerland.

We processed the Iso-Seq data using Pacific Biosciences' SMRTLink platform. Consensus reads were generated from subreads and then demultiplexed. Demultiplexing included removal of barcodes and cDNA primer fragments. During the refine step, polyA tails were trimmed and concatemers removed to produce full-length non-concatemer reads (FLNC). FLNC reads were clustered per sequencing library to produce FASTA files containing assembled transcripts for each sample. Tools used: ccs, 8.2.0; pbtrim, 1.1.0; jasmine, 2.3.0; lima, 2.11.0; skera-split, 1.3.0; lima.1, 2.12.0; refine, 4.2.0; pbmerge, 3.4.0; cluster2, 4.2.0. We then generated gene models and transcript quantification using the OenMel1 reference genome and annotation (*63*) in IsoQuant version 3.6.3 (*89*) the currently most efficient tool for isoform detection (*90*). We used the previously assembled transcripts as input data with options–data_type assembly–fl_data –count_exons. IsoQuant reported an updated gene model containing an additional exon upstream of ASIP that matches existing gene models for ASIP (e.g., chicken in the ENSEMBL gene browser). To improve transcript classification, we re-ran IsoQuant with the updated ASIP gene annotation for each sample. We performed downstream data analysis of IsoQuant results in R.

Evolutionary history of color phenotypes

To reconstruct the evolutionary history of throat and mantle coloration, we first reconstructed the underlying variation's ancestral states. To this end we applied the maximum-likelihood approach implemented in est-sfs (*91*) with *O. deserti*, *O. monacha*, and *O. oenanthe* (*22*) as outgroups and the rate 6 model to estimate the ancestral states in the common ancestor of the *hispanica* complex and *O. deserti*. To infer the longer-term conservation of the two throat SNPs, we furthermore queried the NCBI ortholog database for ASIP, downloaded one amino acid sequence per bird species other than wheatear (N=107) and aligned these along with the two wheatear sequences using NCBI's COBALT multiple alignment tool.

To infer whether similar throat and mantle plumage phenotypes in the *hispanica* complex are based on common variation recruited from ancestral variation, by introgression, or evolved independently from novel variation, we ran topology weighting in TWISST (*92*) under the premise that topologies grouping lineages according to coloration reflect a common genetic basis. TWISST was run reducing the phased dataset to the reference individuals and removing either *O. pleschanka* or the Pontic lineage to only explore three (unrooted) topologies at a time: the species tree [(*O. hispanica, O. pleschanka*/Pontic), *O. melanoleuca, O. cypriaca*)], a coloration tree [(*O. hispanica, O. melanoleuca*), *O. pleschanka*/Pontic, *O. cypriaca*)] grouping white and black lineages, and a topology reflecting known introgression (*26*, *27*) between *O. melanoleuca* and *O. pleschanka*/Pontic [(*O. hispanica*, *O. cypriaca*), (*O. melanoleuca*, *O. pleschanka*/Pontic)]. We generated maximum likelihood trees for TWISST using the phyml_sliding_window.py script (*92*) with the GTR model using 50 SNPs windows in PhyML (*93*). We were interested to understand the extent by which *melanoleuca*- and hybrid-type Pontic (n=11) and Caspian (n=3) individuals were more introgressed by *O. melanoleuca* than parent-type phenotypes in these populations. To this end, we counted the frequency of topologies congruent with color introgression, that is, ones in which *O. melanoleuca* is grouped with the according *melanoleuca*-type/hybrid-type, or with parent-type Pontic or Caspian individuals. We obtained these counts genome-wide, for chromosome 20, and for the 50 kbp ASIP region to understand whether introgression occurred genome-wide or concentrated on chromosome 20 or the ASIP region. To obtain further evidence for or against introgression, we investigated patterns of excessive allele sharing across the ASIP region based on f_{dM}, testing for (i) "white" introgression between *O. hispanica* and *O. melanoleuca*, variably including one of the three black-mantled lineages. As a control comparison and to exclude that introgression occurred between the black-mantled lineages, we estimated f_{dM} between *O. cypriaca, O. pleschanka* and Pontic. We computed the f_{dM} statistic in a sliding window containing 50 SNPs with a step of 5 SNPs using Dsuite (*82*) Dinvestigate and *O. deserti* as outgroup.

To infer the age of the color phenotypes and history of ASIP introgression we compared TMRCAs (logarithmic mean weighted by tree span) for different combinations of species and ASIP haplotypes (tables S3

and S9), with TMRCAs estimated from the ARG using tsinfer and tsdate as outlined above. The time scales at which ASIP haplotypes coalesce relative to interspecific variation provides information on whether color phenotypes are ancestral (interspecific variation coalesces before ASIP haplotypes) or introgressed (interspecific variation coalesces after ASIP haplotypes). We used TMRCAs between species to time the history of species splits within which to place the evolution of ASIP haplotypes, (Fig. 4C and table S3). To estimate the age of black and white throat haplotypes, we identified black-throated and white-throated haplotypes based on which allele was present at the first coding SNP within the ASIP gene that is associated with throat coloration (chromosome 20, position 12,972,212). To estimate the time at which the black and white throat haplotypes split, we estimated the TMRCA within and between all black- and white-throated haplotypes, with haplotypes combined across species. To determine the ancestral haplotype and the origin of the derived haplotype, we further compared TMRCAs within and between black- and white-throated haplotypes for both *melanoleuca* and *hispanica*, identifying the ancestral haplotype as the one coalescing further back in time. Finally, to estimate the timing of introgression, we estimated the TMRCA between white *melanoleuca* and white *hispanica* alleles. The analyses targeting the variation underpinning throat coloration were performed both on the ASIP gene body and on the gene body plus the region upstream to the throat LTR. Analyses to determine the evolutionary history of mantle coloration followed the same logic, grouping haplotypes according to species' mantle coloration, and estimating TMRCAs from the ASIP region upstream the LTR and for the entire 50 kbp comprising the ASIP region. All analyses made use of the reference individuals but excluded five individuals from the reference set for the throat phenotype-based analysis due to poor genotype quality at the diagnostic SNP or throat LTR.

To obtain first insights into the role of selection in fixing ASIP variation associated with alternative coloration, we compared each taxon's unfolded (ancestry-polarized) site frequency spectrum (SFS) of mantle loci against the taxon's SFS with alleles polarized according to the coloration they underpin (Fig. 6A). These comparisons visualize whether the frequencies of alleles underpinning the expressed mantle colors are skewed toward higher frequencies compared to skews in the frequencies in ancestry classes. Under the premise that the coloration each allele underpins rather than its ancestry is exposed to selection, this comparison provides insights into the role of selection in the evolution of species' coloration. Finally, we inferred footprints of positive selection in the ASIP region based on haplotype structure and variation in the TMRCA across the ASIP region. To this end, we inferred XP-nSL (*94*) in selscan v2.0.2 (*95*) and estimated within-taxon TMRCAs as stated above.

Comparative phylogenomics

To infer whether the same SNP variants are responsible for throat and mantle color variation across wheatears and relatives, from 50 taxa and phenotypes we extracted genotypes from mapped reads published in Alaei Kakhki *et al.* (*22*). Based on the same data, we genotyped the throat LTR in each species using Paragraph (*96*) to infer whether the same LTR underpins non-melanic throat coloration in species outside the *hispanica* complex.

Stable isotope analysis

To investigate whether the throat color polymorphism could be associated with differences in dietary composition or environmental variables between phenotypes, we investigated stable isotope signatures in the blood of 161 wheatears. To this end, whole blood preserved in >95% ethanol was used to conduct multi-element stable isotope analysis for the quantification of $\delta^{13}C$, $\delta^{15}N$, and $\delta^{34}S$ values. Dried and powdered samples of approximately 0.3-0.4 mg were weighed to the nearest 0.001 mg in small tin cups on a micro-analytical balance. Samples were then combusted in a vario micro cube elemental analyzer (Elementar Analysensysteme GmbH, Germany). Subsequently, the evolved CO_2, N_2, and SO_2 gases underwent gas chromatographic separation and were introduced into the inlet of an Isoprime isotope ratio mass spectrometer (IRMS) for determination of 13C/12C, 15N/14N ratios and 34S/32S. Measurements are reported in δ-notation ($\delta^{13}C$, $\delta^{15}N$, and $\delta^{34}S$), where

$$\delta_{\text{sample}} = \left[\left(\frac{R_{\text{sample}}}{R_{\text{standard}}}\right) - 1\right] \times 1000$$

relative to standards [Pee Dee Belemnite (PDB) for carbon, atmospheric N_2 for nitrogen, and IAEA-S1 and IAEA-S3 for sulfur] in parts per thousand deviations. Thirty-two sulfanilamide (Iso-prime internal) standards were placed between five unknowns in sequence. These Internal laboratory standards indicated measurement errors (SD) of ± 0.08‰ for $\delta^{13}C$, 0.12‰ for $\delta^{15}N$ and 0.23‰ for $\delta^{34}S$. To infer whether throat plumage phenotypes differed in blood isotope composition, we ran linear models with log-transformed $\delta^{13}C$ as response variables, throat coloration as explanatory variable, and coordinates (latitude, longitude), birds' age, day in the year, and year as covariables. Interactions between latitude and longitude, and day in the year and year were included in the models. We performed these analyses on the complete data set (n=161) as well as separately for *melanoleuca* individuals (n=92), Greek *melanoleuca* individuals (n=57), and Iranian individuals (n=69) to account for potential large-scale geographic isotope variation not captured by coordinates. To account for potential biases connected with the smaller sample size of white-throated individuals, we performed a resampling procedure. To this end, we repeated the aforementioned analyses by resampling (10^5 times) as many black-throated individuals as white-throated individuals were available for the respective data partition. To understand whether the mean and variance of $\delta^{15}N$ values observed for the white individuals differ from the ones of the same sample sizes of black-throated individuals, we plotted the mean and variance of the $\delta^{15}N$ values observed for white-throated individuals against the distributions of the according resampled values for black-throated individuals (fig. S16, left and middle). To confirm that these results also hold after accounting for covariates, for each of the 10^5 samples we furthermore recorded the effect size of throat coloration from a linear model including the same covariates as the linear model run on the full data. Based on this, we tested whether the distributions of effect sizes of throat coloration adjusted for sample size and covariates differ from zero (fig. S16, right).

REFERENCES AND NOTES

1. M. Nei, The new mutation theory of phenotypic evolution. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12235–12242 (2007). doi: 10.1073/pnas.0703349104; pmid: 17640887
2. K. F. Stryjewski, M. D. Sorenson, Mosaic genome evolution in a recent and rapid avian radiation. *Nat. Ecol. Evol.* **1**, 1912–1922 (2017). doi: 10.1038/s41559-017-0364-7; pmid: 29085063
3. S. Lamichhaney *et al.*, A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science* **352**, 470–474 (2016). doi: 10.1126/science.aad8786; pmid: 27102486
4. Heliconius Genome Consortium, Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**, 94–98 (2012). doi: 10.1038/nature11041; pmid: 22722851
5. S. P. Turbek *et al.*, Rapid speciation via the evolution of pre-mating isolation in the Iberá Seedeater. *Science* **371**, eabc0256 (2021). doi: 10.1126/science.abc0256; pmid: 33766854
6. D. A. Marques, J. I. Meier, O. Seehausen, A Combinatorial View on Speciation and Adaptive Radiation. *Trends Ecol. Evol.* **34**, 531–544 (2019). doi: 10.1016/j.tree.2019.02.008; pmid: 30885412
7. T. M. Anderson *et al.*, Molecular and evolutionary history of melanism in North American gray wolves. *Science* **323**, 1339–1343 (2009). doi: 10.1126/science.1165448; pmid: 19197024
8. E. Huerta-Sánchez *et al.*, Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **512**, 194–197 (2014). doi: 10.1038/nature13408; pmid: 25043035

9. R. W. R. Wallbank *et al.*, Evolutionary Novelty in a Butterfly Wing Pattern through Enhancer Shuffling. *PLOS Biol.* **14**, e1002353 (2016). doi: 10.1371/journal.pbio.1002353; pmid: 26771987
10. P. F. Colosimo *et al.*, Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* **307**, 1928–1933 (2005). doi: 10.1126/science.1107239; pmid: 15790847
11. S. Wang *et al.*, Selection on a small genomic region underpins differentiation in multiple color traits between two warbler species. *Evol. Lett.* **4**, 502–515 (2020). doi: 10.1002/evl3.198; pmid: 33312686
12. M. D. Baiz, A. W. Wood, A. Brelsford, I. J. Lovette, D. P. L. Toews, Pigmentation Genes Show Evidence of Repeated Divergence and Multiple Bouts of Introgression in *Setophaga* Warblers. *Curr. Biol.* **31**, 643–649.e3 (2021). doi: 10.1016/j.cub.2020.10.094; pmid: 33259789
13. J. I. Meier *et al.*, Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* **8**, 14363 (2017). doi: 10.1038/ncomms14363; pmid: 28186104
14. J. I. Meier *et al.*, Cycles of fusion and fission enabled rapid parallel adaptive radiations in African cichlids. *Science* **381**, eade2833 (2023). doi: 10.1126/science.ade2833; pmid: 37769075
15. L. Campagna *et al.*, Repeated divergent selection on pigmentation genes in a rapid finch radiation. *Sci. Adv.* **3**, e1602404 (2017). doi: 10.1126/sciadv.1602404; pmid: 28560331
16. B. Prud'homme *et al.*, Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature* **440**, 1050–1053 (2006). doi: 10.1038/nature04597; pmid: 16625197
17. D. L. Bannasch *et al.*, Dog colour patterns explained by modular promoters of ancient canid origin. *Nat. Ecol. Evol.* **5**, 1415–1423 (2021). doi: 10.1038/s41559-021-01524-x; pmid: 34385618
18. M. Gautier *et al.*, The Genomic Basis of Color Pattern Polymorphism in the Harlequin Ladybird. *Curr. Biol.* **28**, 3296–3302.e7 (2018). doi: 10.1016/j.cub.2018.08.023; pmid: 30146156
19. M. Bohutínská, C. L. Peichel, Divergence time shapes gene reuse during repeated adaptation. *Trends Ecol. Evol.* (2023). doi: 10.1016/j.tree.2023.11.007; pmid: 38155043
20. J. Arendt, D. Reznick, Convergence and parallelism reconsidered: What have we learned about the genetics of adaptation? *Trends Ecol. Evol.* **23**, 26–32 (2008). doi: 10.1016/j.tree.2007.09.011; pmid: 18022278
21. J. Cerca, Understanding natural selection and similarity: Convergent, parallel and repeated evolution. *Mol. Ecol.* **32**, 5451–5462 (2023). doi: 10.1111/mec.17132; pmid: 37724599
22. N. Alaei Kakhki *et al.*, A Phylogenomic Assessment of Processes Underpinning Convergent Evolution in Open-Habitat Chats. *Mol. Biol. Evol.* **40**, msac278 (2023). doi: 10.1093/molbev/msac278; pmid: 36578177
23. M. Aliabadian *et al.*, Convergent evolution of morphological and ecological traits in the open-habitat chat complex (Aves, Muscicapidae: Saxicolinae). *Mol. Phylogenet. Evol.* **65**, 35–45 (2012). doi: 10.1016/j.ympev.2012.05.011; pmid: 22634240
24. M. Schweizer *et al.*, Genome-wide evidence supports mitochondrial relationships and pervasive parallel phenotypic evolution in open-habitat chats. *Mol. Phylogenet. Evol.* **139**, 106568 (2019). doi: 10.1016/j.ympev.2019.106568; pmid: 31349102
25. M. Schweizer *et al.*, Parallel plumage colour evolution and introgressive hybridization in wheatears. *J. Evol. Biol.* **32**, 100–110 (2019). doi: 10.1111/jeb.13401; pmid: 30421480
26. J. Haffer, Secondary contact zones of birds in Northern Iran. *Bonn. Zool. Monogr.* **10**, 1–64 (1977).
27. E. N. Panov, *Wheaters of the Palearctic. Ecology, Behaviour and Evolution of the genus Oenanthe* (Pensoft, 2005).
28. D. J. Lawson, G. Hellenthal, S. Myers, D. Falush, Inference of population structure using dense haplotype data. *PLOS Genet.* **8**, e1002453 (2012). doi: 10.1371/journal.pgen.1002453; pmid: 22291602
29. C. A. Buerkle, C. Lexer, Admixture as the basis for genetic mapping. *Trends Ecol. Evol.* **23**, 686–694 (2008). doi: 10.1016/j.tree.2008.07.008; pmid: 18845358
30. X. Zhou, M. Stephens, Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012). doi: 10.1038/ng.2310; pmid: 22706312
31. H. E. Hoekstra, Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity* **97**, 222–234 (2006). doi: 10.1038/sj.hdy.6800861; pmid: 16823403
32. G. A. Semenov *et al.*, Asymmetric introgression reveals the genetic architecture of a plumage trait. *Nat. Commun.* **12**, 1019 (2021). doi: 10.1038/s41467-021-21340-y; pmid: 33589637
33. S. J. Lin *et al.*, Topology of feather melanocyte progenitor niche allows complex pigment patterns to emerge. *Science* **340**, 1442–1445 (2013). doi: 10.1126/science.1230374; pmid: 23618762
34. C. R. Linnen *et al.*, Adaptive evolution of multiple traits through multiple mutations at a single gene. *Science* **339**, 1312–1316 (2013). doi: 10.1126/science.1233213; pmid: 23493712
35. N. Haupaix *et al.*, The periodic coloration in birds forms through a prepattern of somite origin. *Science* **361**, eaar4777 (2018). doi: 10.1126/science.aar4777; pmid: 30237324
36. D. Liang *et al.*, Genomic Analysis Revealed a Convergent Evolution of LINE-1 in Coat Color: A Case Study in Water Buffaloes (*Bubalus bubalis*). *Mol. Biol. Evol.* **38**, 1122–1136 (2021). doi: 10.1093/molbev/msaa279; pmid: 33212507
37. J. Kelleher *et al.*, Inferring whole-genome histories in large population datasets. *Nat. Genet.* **51**, 1330–1338 (2019). doi: 10.1038/s41588-019-0483-y; pmid: 31477934
38. A. W. Wohns *et al.*, A unified genealogy of modern and ancient genomes. *Science* **375**, eabi8264 (2022). doi: 10.1126/science.abi8264; pmid: 35201891
39. R. Inger, S. Bearhop, Applications of stable isotope analyses to avian ecology. *Ibis* **150**, 447–461 (2008). doi: 10.1111/j.1474-919X.2008.00839.x
40. K. A. Hobson, Tracing origins and migration of wildlife using stable isotopes: A review. *Oecologia* **120**, 314–326 (1999). doi: 10.1007/s004420050865; pmid: 28308009
41. E. Yohannes, J.-L. Berthoud, F. Woog, Trait based niche differentiation in tetrakas (Bernieridae) endemic to Madagascar: A multi-isotope approach. *Front. Ecol. Evol.* **11**, 1082226 (2023). doi: 10.3389/fevo.2023.1082226
42. L. M. San-Jose *et al.*, Differential fitness effects of moonlight on plumage colour morphs in barn owls. *Nat. Ecol. Evol.* **3**, 1331–1340 (2019). doi: 10.1038/s41559-019-0967-2; pmid: 31477846
43. R. L. Mumme, White tail spots and tail-flicking behavior enhance foraging performance in the Hooded Warbler. *Auk* **131**, 141–149 (2014). doi: 10.1642/AUK-13-199.1
44. P. G. Jablonski, A rare predator exploits prey escape behavior: The role of tail-fanning and plumage contrast in foraging of the painted redstart (*Myioborus pictus*). *Behav. Ecol.* **10**, 7–14 (1999). doi: 10.1093/beheco/10.1.7
45. A.-L. Ducrest, L. Keller, A. Roulin, Pleiotropy in the melanocortin system, coloration and behavioural syndromes. *Trends Ecol. Evol.* **23**, 502–510 (2008). doi: 10.1016/j.tree.2008.06.001; pmid: 18644658
46. N. Rosser *et al.*, Hybrid speciation driven by multilocus introgression of ecological traits. *Nature* **628**, 811–817 (2024). doi: 10.1038/s41586-024-07263-w; pmid: 38632397
47. C. Estalles *et al.*, Concerted variation in melanogenesis genes underlies emergent patterning of plumage in capuchino seedeaters. *Proc. Biol. Sci.* **289**, 20212277 (2022). doi: 10.1098/rspb.2021.2277; pmid: 35016545
48. D. P. L. Toews *et al.*, Plumage Genes and Little Else Distinguish the Genomes of Hybridizing Warblers. *Curr. Biol.* **26**, 2313–2318 (2016). doi: 10.1016/j.cub.2016.06.034; pmid: 27546575
49. S. Y. W. Sin *et al.*, Genetic basis and evolution of structural color polymorphism in an Australian songbird. *Mol. Biol. Evol.* **41**, msae046 (2024). doi: 10.1093/molbev/msae046; pmid: 38415852
50. H. Vrieling, D. M. Duhl, S. E. Millar, K. A. Miller, G. S. Barsh, Differences in dorsal and ventral pigmentation result from regional expression of the mouse agouti gene. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5667–5671 (1994). doi: 10.1073/pnas.91.12.5667; pmid: 8202545
51. J. D. Galbraith, A. Hayward, The influence of transposable elements on animal colouration. *Trends Genet.* **39**, 624–638 (2023). doi: 10.1016/j.tig.2023.04.005; pmid: 37183153
52. M. Rebeiz, J. E. Pool, V. A. Kassner, C. F. Aquadro, S. B. Carroll, Stepwise modification of a modular enhancer underlies adaptation in a *Drosophila* population. *Science* **326**, 1663–1667 (2009). doi: 10.1126/science.1178357; pmid: 20019281
53. R. D. H. Barrett, D. Schluter, Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23**, 38–44 (2008). doi: 10.1016/j.tree.2007.09.008; pmid: 18006185
54. H. E. Hoekstra, J. A. Coyne, The locus of evolution: Evo devo and the genetics of adaptation. *Evolution* **61**, 995–1016 (2007). doi: 10.1111/j.1558-5646.2007.00105.x; pmid: 17492956
55. S. B. Carroll, Endless forms: The evolution of gene regulation and morphological diversity. *Cell* **101**, 577–580 (2000). doi: 10.1016/S0092-8674(00)80868-5; pmid: 10892643
56. C. Mérot, R. A. Oomen, A. Tigano, M. Wellenreuther, A Roadmap for Understanding the Evolutionary Significance of Structural Genomic Variation. *Trends Ecol. Evol.* **35**, 561–572 (2020). doi: 10.1016/j.tree.2020.03.002; pmid: 32521241
57. A. Eyre-Walker, P. D. Keightley, The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* **8**, 610–618 (2007). doi: 10.1038/nrg2146; pmid: 17637733
58. E. B. Chuong, N. C. Elde, C. Feschotte, Regulatory activities of transposable elements: From conflicts to benefits. *Nat. Rev. Genet.* **18**, 71–86 (2017). doi: 10.1038/nrg.2016.139; pmid: 27867194
59. J. V. Peñalba *et al.*, The Role of Hybridization in Species Formation and Persistence. *Cold Spring Harb. Perspect. Biol.* **16**, a041445 (2024). doi: 10.1101/cshperspect.a041445; pmid: 38438186
60. D. Lutgen *et al.*, Linked-read sequencing enables haplotype-resolved resequencing at population scale. *Mol. Ecol. Resour.* **20**, 1311–1322 (2020). doi: 10.1111/1755-0998.13192; pmid: 32419391
61. M. Baym *et al.*, Inexpensive multiplexed library preparation for megabase-sized genomes. *PLOS ONE* **10**, e0128036 (2015). doi: 10.1371/journal.pone.0128036; pmid: 26000737
62. A. McKenna *et al.*, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010). doi: 10.1101/gr.107524.110; pmid: 20644199
63. V. Peona *et al.*, An annotated chromosome-scale reference genome for Eastern Black-eared Wheatear (Oenanthe melanoleuca). *G3* **17**, jkad088 (2023).
64. A. Bishara *et al.*, Read clouds uncover variation in complex regions of the human genome. *Genome Res.* **25**, 1570–1580 (2015). doi: 10.1101/gr.191189.115; pmid: 26286554
65. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018). doi: 10.1093/bioinformatics/bty560; pmid: 30423086

66. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009). doi: 10.1093/bioinformatics/btp324; pmid: 19451168
67. P. Edge, V. Bafna, V. Bansal, HapCUT2: Robust and accurate haplotype assembly for diverse sequencing technologies. *Genome Res.* **27**, 801–812 (2017). doi: 10.1101/gr.213462.116; pmid: 27940952
68. M. Martin, M. Patterson, S. Garg, S. O. Fischer, N. Pisanti, G. W. Klau, A. Schöenhuth, T. Marschall, WhatsHap: fast and accurate read-based phasing. bioRxiv 085050 [Preprint] (2016); https://doi.org/10.1101/085050.
69. O. Delaneau, J.-F. Zagury, M. R. Robinson, J. L. Marchini, E. T. Dermitzakis, Accurate, scalable and integrative haplotype estimation. *Nat. Commun.* **10**, 5436 (2019). doi: 10.1038/s41467-019-13225-y; pmid: 31780650
70. N. Patterson, A. L. Price, D. Reich, Population structure and eigenanalysis. *PLOS Genet.* **2**, e190 (2006). doi: 10.1371/journal.pgen.0020190; pmid: 17194218
71. S. Purcell *et al.*, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007). doi: 10.1086/519795; pmid: 17701901
72. P. Danecek *et al.*, Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021). doi: 10.1093/gigascience/giab008; pmid: 33590861
73. E. P. Derryberry, G. E. Derryberry, J. M. Maley, R. T. Brumfield, HZAR: Hybrid zone analysis using an R software package. *Mol. Ecol. Resour.* **14**, 652–663 (2014). doi: 10.1111/1755-0998.12209; pmid: 24373504
74. Z. Gompert, D. A. DeRaad, C. A. Buerkle, A Next Generation of Hierarchical Bayesian Analyses of Hybrid Zones Enables Model-Based Quantification of Variation in Introgression in R. *Ecol. Evol.* **14**, e70548 (2024). doi: 10.1002/ece3.70548; pmid: 39583044
75. J. P. Spence, Y. S. Song, Inference and analysis of population-specific fine-scale recombination maps across 26 diverse human populations. *Sci. Adv.* **5**, eaaw9206 (2019). doi: 10.1126/sciadv.aaw9206; pmid: 31681842
76. L. Smeds, A. Qvarnström, H. Ellegren, Direct estimate of the rate of germline mutation in a bird. *Genome Res.* **26**, 1211–1218 (2016). doi: 10.1101/gr.204669.116; pmid: 27412854
77. T. C. Bruen, H. Philippe, D. Bryant, A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**, 2665–2681 (2006). doi: 10.1534/genetics.105.048975; pmid: 16489234
78. H. Ellegren *et al.*, The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* **491**, 756–760 (2012). doi: 10.1038/nature11584; pmid: 23103876
79. C. Zhang, M. Rabiee, E. Sayyari, S. Mirarab, ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19** (Suppl 6), 153 (2018). doi: 10.1186/s12859-018-2129-y; pmid: 29745866
80. B. Q. Minh *et al.*, IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020). doi: 10.1093/molbev/msaa015; pmid: 32011700
81. S. Schiffels, R. Durbin, Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* **46**, 919–925 (2014). doi: 10.1038/ng.3015; pmid: 24952747
82. M. Malinsky, M. Matschiner, H. Svardal, Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Mol. Ecol. Resour.* **21**, 584–595 (2021). doi: 10.1111/1755-0998.13265; pmid: 33012121
83. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014). doi: 10.1186/s12859-014-0356-4; pmid: 25420514
84. C. Goubert *et al.*, A beginner's guide to manual curation of transposable elements. *Mob. DNA* **13**, 7 (2022). doi: 10.1186/s13100-021-00259-7; pmid: 35354491
85. J. R. Belyeu *et al.*, Samplot: A platform for structural variant visual validation and automated filtering. *Genome Biol.* **22**, 161 (2021). doi: 10.1186/s13059-021-02380-5; pmid: 34034781
86. J. R. Belyeu *et al.*, SV-plaudit: A cloud-based framework for manually curating thousands of structural variants. *Gigascience* **7**, giy064 (2018). doi: 10.1093/gigascience/giy064; pmid: 29860504
87. P.-C. Bürkner, brms: An R Package for Bayesian Multilevel Models Using Stan. *J. Stat. Softw.* **80**, (2017). doi: 10.18637/jss.v080.i01
88. A. Vehtari, A. Gelman, J. Gabry, Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432 (2017). doi: 10.1007/s11222-016-9696-4
89. A. D. Prjibelski *et al.*, Accurate isoform discovery with IsoQuant using long reads. *Nat. Biotechnol.* **41**, 915–918 (2023). doi: 10.1038/s41587-022-01565-y; pmid: 36593406
90. Y. Su *et al.*, Comprehensive assessment of mRNA isoform detection methods for long-read sequencing data. *Nat. Commun.* **15**, 3972 (2024). doi: 10.1038/s41467-024-48117-3; pmid: 38730241
91. P. D. Keightley, B. C. Jackson, Inferring the Probability of the Derived *vs.* the Ancestral Allelic State at a Polymorphic Site. *Genetics* **209**, 897–906 (2018). doi: 10.1534/genetics.118.301120; pmid: 29769282
92. S. H. Martin, S. M. Van Belleghem, Exploring Evolutionary Relationships Across the Genome Using Topology Weighting. *Genetics* **206**, 429–438 (2017). doi: 10.1534/genetics.116.194720; pmid: 28341652
93. S. Guindon *et al.*, New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010). doi: 10.1093/sysbio/syq010; pmid: 20525638
94. Z. A. Szpiech, T. E. Novak, N. P. Bailey, L. S. Stevison, Application of a novel haplotype-based scan for local adaptation to study high-altitude adaptation in rhesus macaques. *Evol. Lett.* **5**, 408–421 (2021). doi: 10.1002/evl3.232; pmid: 34367665
95. Z. A. Szpiech, selscan 2.0: Scanning for sweeps in unphased data. *Bioinformatics* **40**, btae006 (2024). doi: 10.1093/bioinformatics/btae006; pmid: 38180866
96. S. Chen *et al.*, Paragraph: A graph-based structural variant genotyper for short-read sequence data. *Genome Biol.* **20**, 291 (2019). doi: 10.1186/s13059-019-1909-7; pmid: 31856913
97. D. Lutgen *et al.*, Supporting material for: A mosaic of modular variation at a single gene underpins convergent plumage coloration, Zenodo (2025); https://doi.org/10.5281/zenodo.15487817.
98. F. Suárez, El dimorfismo de la collalba rubia (Oenanthe hispanica L.): Variaciones geograficas y entre habitats. *Ardeola* **37**, 291–298 (1990).
99. BirdLife International, *Handbook of the Birds of the World, Bird species distribution maps of the world, version 6.0* (Lynx, 2016).

ACKNOWLEDGMENTS

We thank A. Khani, R. Ramezani, M. Monfared, S. Huang, V. Khimonen, C. Nikiforou, M. Dretakis, S. Iezekiel, C. Grande, M. Illa, M. Pérez, M. Dimaki, C. Ninou, and J. Rodriguez Fuentes for support with fieldwork and M. Alaei Kakhki; the regional water company of Khorasan Razavi, Northern Khorasan, Semnan, Ardabil, and Tabriz; Ferdowsi University of Mashhad; I. Grigoryeva; V. Voronova; and F. Jiguet for logistic support. We thank P. Korner for statistics support, A. Gomez for support with DNA extractions, S. Schiffels for support with data analysis, and the Peichel group, the Suh group, and Jochen Wolf for discussion. M. Carneiro and S. Yin provided guidance on feather sampling for transcriptome analysis. We thank the University of Erlangen, Applied and Environmental Geology, and the University of Graz, Department of Earth Sciences, for the support in the triple-element stable isotope measurements. We thank all permit-issuing authorities. The Burke Museum of the University of Washington, the Museum of Comparative Zoology of Harvard University, the Steinhardt Museum of Natural History in Tel Aviv, the Royal Ontario Museum in Toronto, and Hadoram Shirihai kindly provided samples. Computations were performed at the Genetic Diversity Centre (GDC), ETH Zurich, the High-Performance Computing Cluster EVE, a joint effort of the Helmholtz Centre for Environmental Research (UFZ) and the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig; at UPPMAX with resources provided by the project NAISS 2024/22-351 from the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by the Swedish Research Council through grant agreement no. 2022-06725; and at the University of Cyprus High Performance Cluster. We thank the administration and support staff of GDC, EVE, and NAISS. We thank the sequencing facilities of NGI Sweden in Solna and the NGS platform of the University of Berne and their staff for their support. **Funding:** This research was supported by German Research Foundation (DFG) research grant BU3456/3-1 and Swiss National Science Foundation (SNSF) grant 310030_207417 to R.B., National Research Fund (FNR) Luxembourg grant 14575729 to D.L., an A.G. Leventis Foundation Grant to A.N.G.K., a Georg Foster Research Stipend of the Alexander von Humboldt Foundation and a scholarship for female researchers from Friedrich-Schiller-University Jena to N.A.K., and a Swedish Research Council (Vetenskapsrådet) grant 2022-06195 to V.P. **Author contributions:** R.B. conceptualized the study with input from D.L., A.N.G.K., M.S., and H.S. D.L., V.P., M.A.C., N.A.K., F.T.-L., E.Y., and S.G.d.S. performed analyses. P.-A.G. supported data analysis. D.L., A.-L.D., M.B., and A.U. performed lab work. R.B., M.S., H.S., P.-A.G., A.N.G.K., A.S., and C.L.P. supervised analyses. P.A., S.M.L., M.M., E.Y., A.A., T.A., M.A., N.A., V.B., I.C., J.L.C., E.F., J.T.G., Z.J., A.K., F.T.-L., Y.L., N.P., R.P., Á.P., R.R., A.D.S., F.S., P.S., S.S., B.S., A.T., N.V., F.T.-L., J.G., and L.Y. provided resources. R.B. and D.L. wrote the paper with input from M.A.C., P.-A.G., A.N.G.K., F.T.-L., C.L.P., V.P., H.S., M.S., and A.S. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Sequencing data and the linked read-based *vittata* genome assembly are available on the short-read archive, bioprojects PRJNA1128009 (whole-genome resequencing data) and PRJNA1260766 (Iso-Seq data). Stable isotope data and scripts used for analyses are available on Zenodo (*97*). **License information:** Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ado8005
Supplementary Text; Figs. S1 to S25; Tables S1 to S17; References (*100*); Data S1 to S3; MDAR Reproducibility Checklist

Submitted 24 May 2024; resubmitted 24 March 2025; accepted 17 July 2025
10.1126/science.ado8005

RESEARCH ARTICLE SUMMARY

IMAGING

Mesoscale volumetric fluorescence imaging at nanoscale resolution by photochemical sectioning

Wei Wang†, Xiongtao Ruan†, Gaoxiang Liu†, Daniel E. Milkie, Wenping Li, Eric Betzig, Srigokul Upadhyayula*, Ruixuan Gao*

Full article and list of author affiliations: https://doi.org/10.1126/science.adr9109

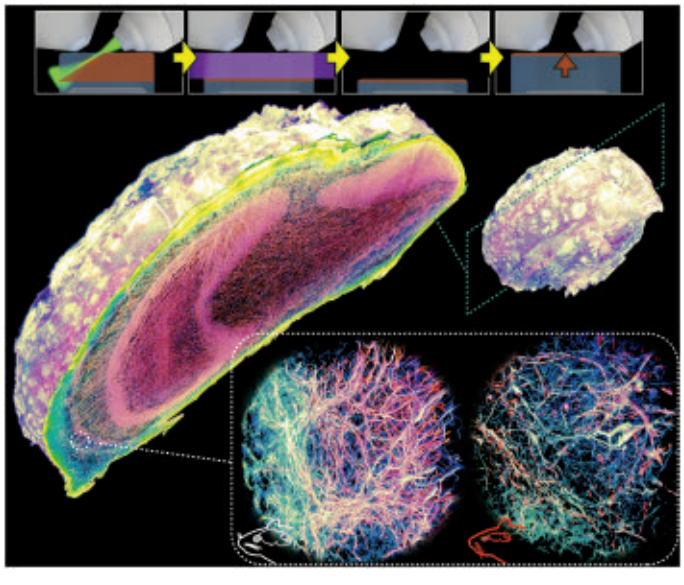
INTRODUCTION: Imaging biological structures at nanoscale resolution deep into large specimens is constrained by the short working distance of objectives used in high-resolution optical microscopes. Hydrogel-based clearing and expansion, while enhancing optical transparency and effective resolution, further push large biological specimens beyond the objective's reach. Mechanical sectioning can circumvent this limit but often introduces distortion, tearing, and loss that complicate large-scale, high-fidelity reconstructions, particularly in hydrogel-embedded and expanded specimens. These limits preclude continuous nanoscale imaging across an intact, whole-mount specimen such as a mammalian brain.

RATIONALE: We sought a contact-free, spatially precise alternative to mechanical sectioning that enables nanoscale imaging of whole-mount specimens beyond the microscope's working distance. For this purpose, we combined a photochemical strategy for sample sectioning and sequential on-block optical imaging in a workflow we call VIPS (volumetric imaging via photochemical sectioning). After each imaging cycle, a confined illumination removes a portion of the light-sensitive, hydrogel-embedded specimen to enable a new round of on-block imaging. Iterating these steps, layer by layer, to the desired depth preserves molecular contrast from the anchored biomolecules and yields contiguous volumetric datasets.

RESULTS: To implement VIPS, we developed a superabsorbent hydrogel containing a photocleavable cross-linker responsive to ultraviolet excitation. Biological specimens embedded and expanded in this hydrogel underwent rapid, complete photodegradation under either single- or multiphoton illumination. Alternating cycles of on-block lattice light-sheet microscopy (LLSM) and light-sheet photochemical sectioning allowed mesoscale volumetric imaging at nanoscale resolution across two complete mouse olfactory bulbs, one wild-type and one neurodegenerative. A high-performance, petabyte-scale computation pipeline stitched contiguous volumes; reconstructed myelinated and unmyelinated axons; and quantified axon density, tractography, and myelination profiles throughout the bulbs. Olfactory bulb–wide analysis revealed anatomical heterogeneities undetectable in smaller volumes, including a centripetal degeneration pattern of axons in the neurodegenerative bulb.

CONCLUSION: VIPS overcomes working-distance limits by coupling iterative photochemical sectioning with high-resolution optical imaging. This noncontact process is compatible with various hydrogel-based clearing and expansion chemistries and circumvents the distortion, sample loss, and protocol limitations inherent to mechanical sectioning. The method integrates readily into existing microscopes for continuous, automated acquisition, enabling nanoscale fluorescence imaging of whole-mount specimens of virtually unlimited size. Together with LLSM and scalable computation, VIPS effectively removes physical sample size as a practical barrier to super-resolution volumetric imaging, allowing for cellular and subcellular studies of intact specimens at petabyte scale and beyond. Remaining challenges include labeling depth and specificity, data storage and computation, and the need for AI systems to interpret massive datasets. Even so, VIPS advances the field from sparse sampling toward quantitative measurements of biological stereotypy and variability in whole specimens. □

*Corresponding author. Email: gaor@uic.edu (R.G.); sup@berkeley.edu (S.U.) The list of author affiliations is available in the full article online. †These authors contributed equally to this work.
Cite this article as W. Wang *et al.*, *Science* **390**, eadr9109 (2025). DOI: 10.1126/science.adr9109



VIPS overview. VIPS overcomes working-distance limits by light-driven photochemical specimen removal (top). Iterative on-block imaging (imaged volume, dark orange) interleaved with photochemical sectioning produces contiguous volumes, effectively eliminating physical sample size as a constraint. This enables super-resolution imaging of entire mouse olfactory bulbs (center) and comparative, whole-mount analysis of axons and myelin sheaths in wild-type versus neurodegenerative mice (bottom insets).

IMAGING

Mesoscale volumetric fluorescence imaging at nanoscale resolution by photochemical sectioning

Wei Wang[1]†, Xiongtao Ruan[2]†, Gaoxiang Liu[2]†, Daniel E. Milkie[3], Wenping Li[1], Eric Betzig[2,3,4], Srigokul Upadhyayula[2,5,6]*, Ruixuan Gao[1,7]*

Optical nanoscopy of intact biological specimens has been transformed by recent advancements in hydrogel-based tissue clearing and expansion, enabling the imaging of cellular and subcellular structures with molecular contrast. However, existing high-resolution fluorescence microscopes are physically limited by objective-to-specimen distance, which prevents the study of whole-mount specimens without physical sectioning. To address this challenge, we developed a photochemical strategy for spatially precise sectioning of specimens. By combining serial photochemical sectioning with lattice light-sheet imaging and petabyte-scale computation, we imaged and reconstructed axons and myelin sheaths across entire mouse olfactory bulbs at nanoscale resolution. An olfactory bulb–wide analysis of myelinated and unmyelinated axons revealed distinctive patterns of axon degeneration and de-/dysmyelination in the neurodegenerative brain, highlighting the potential for peta- to exabyte-scale super-resolution studies using this approach.

Attempts to understand human cognition, behavior, and consciousness have catalyzed new technologies to uncover the anatomical complexity and molecular workings of the brain (*1*, *2*). Among these, tissue clearing and expansion microscopy (ExM) have emerged as effective tools for optical investigation of intact brain tissue (*3*–*6*). In particular, hydrogel-based tissue clearing and expansion has enabled volumetric mapping of biomolecules down to nanoscale resolution through conventional fluorescence microscopy while simultaneously minimizing optical scattering and aberration (*5*–*8*). Tissue expansion plus light-sheet fluorescence microscopy has created a scalable platform with which to study synaptic ultrastructure, connectivity, and molecular compositions of invertebrate and rodent brains with high speed, low photobleaching, and three-dimensional (3D) nanoscale resolution (*9*–*12*).

In ExM, both the expansion factor and the microscope parameters contribute to the overall resolution and scalability. Optimizing both is crucial for achieving isotropic nanoscale resolution, high sensitivity, low photobleaching, and fast imaging across large, expanded samples. In expansion lattice light-sheet microscopy (ExLLSM), for example, although the imaging speed exceeds 100 million voxels per second with minimal photobleaching and sub–100-nm resolution at a modest ~4-fold expansion (*9*), the imaging depth is physically limited by the short working distance of the high numerical aperture (NA) objectives used (*13*). More recent studies have adopted ultralong-working-distance objectives for large-scale imaging (*14*), but often at the expense of NA, and thus resolution, for a given expansion factor. Nevertheless, balancing NA and working distance continues to pose an obstacle in scaling up ExM-based methods to whole-mount specimens without sample sectioning.

Electron microscopy (EM) has long faced similar challenges of limited penetration in volumetric imaging. In response, techniques such as serial section electron microscopy and focused ion beam scanning electron microscopy (FIB-SEM) use mechanical sectioning or surface ablation followed by on-section or block-face imaging (Fig. 1A) (*15*–*18*). Although analogous serial sectioning has been applied to volumetric fluorescence imaging of fixed tissue (*19*–*25*), mechanical sectioning of hydrogel-expanded specimens often leads to sample distortion, tearing, and loss (figs. S1 and S2 and movies S1 and S2). Hydrogel stiffening (e.g., increasing monomer concentration, re-embedding, or modifying monomer chemistry) can reduce these problems (*12*, *26*), but the resulting changes in polymer density and composition often lead to reduced imaging resolution (fig. S3) and compromised compatibility with established expansion protocols that require specific hydrogel densities and compositions. Furthermore, in applications such as brain connectomics, sectioning must be nearly perfect because even a single tear or lost section can be catastrophic for tracing and reconstruction of thin processes (*27*, *28*). However, standard vibratomes and (ultra) microtomes struggle to accommodate large mammalian tissues, especially after expansion (*28*).

To overcome these challenges, we developed VIPS (volumetric imaging via photochemical sectioning), a method that extends nanoscale fluorescence imaging to centimeter-scale volumes and beyond. In analogy to FIB-SEM, VIPS uses a spatially precise photochemical degradation process to perform sequential sample degradation and on-block volumetric optical imaging of hydrogel-embedded biological specimens (Fig. 1A and Movie 1). This method uses a superabsorbent hydrogel that incorporates a photocleavable cross-linker to rapidly and completely disintegrate under either ultraviolet (UV) illumination or multiphoton excitation. By combining sequential on-block volumetric lattice light-sheet microscopy (LLSM) imaging, light-sheet photochemical sectioning (PS), and petabyte-scale computation, we imaged and reconstructed axonal projections and myelin sheaths across two complete mouse olfactory bulbs (OBs) at nanoscale resolution. Finally, we performed a comparative analysis of the wild-type (WT) and neurodegenerative mouse OBs, which revealed unique spatial patterns of axon degeneration and de-/dysmyelination in the neurodegenerative sample at single-axon resolution.

Design and synthesis of a photodegradable superabsorbent hydrogel

Photoresponsive hydrogels have been used to modulate mechanical and chemical environments around encapsulated cells (*29*–*32*). After evaluating various photochemistries for their optical properties and compatibilities with tissue expansion and clearing, we selected a polyacrylamide–sodium polyacrylate hydrogel for VIPS. We hypothesized that this polymer, which is widely used for tissue clearing and expansion, could undergo swift and complete photodegradation by incorporating a photocleavable cross-linker (PC) into its polymer network (Fig. 1B). To enhance the usability and accessibility of VIPS, we prioritized the following design characteristics for PC: (i) photocleavage at wavelengths distinct from popular fluorophores; (ii) bio-orthogonal and water-soluble; (iii) low molecular weight for enhanced sample permeability; (iv) compatible with established hydrogel recipes by simply replacing the standard cross-linker; and (v) synthesizable in a minimal number of steps with common wet-lab equipment.

Following these principles, we designed the molecular structure and synthetic pathway of PC (Fig. 1C). This PC molecule, equipped with a ~2-kDa polyethylene glycol backbone and two photocleavable *o*-nitrobenzyl ester moieties, can be synthesized in the ambient environment with standard glassware and wet-lab equipment (see the materials and methods and fig. S4). Using PC as the cross-linker, we then

[1]Department of Chemistry, University of Illinois Chicago, Chicago, IL, USA. [2]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA, USA. [3]Howard Hughes Medical Institute, Janelia Research Campus, Ashburn, VA, USA. [4]Department of Physics, Howard Hughes Medical Institute, Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA, USA. [5]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [6]Chan Zuckerberg Biohub, San Francisco, CA, USA. [7]Department of Biological Sciences, University of Illinois Chicago, Chicago, IL, USA. ***Corresponding author. Email: gaor@uic.edu (R.G.); sup@berkeley.edu (S.U.)** †These authors contributed equally to this work.

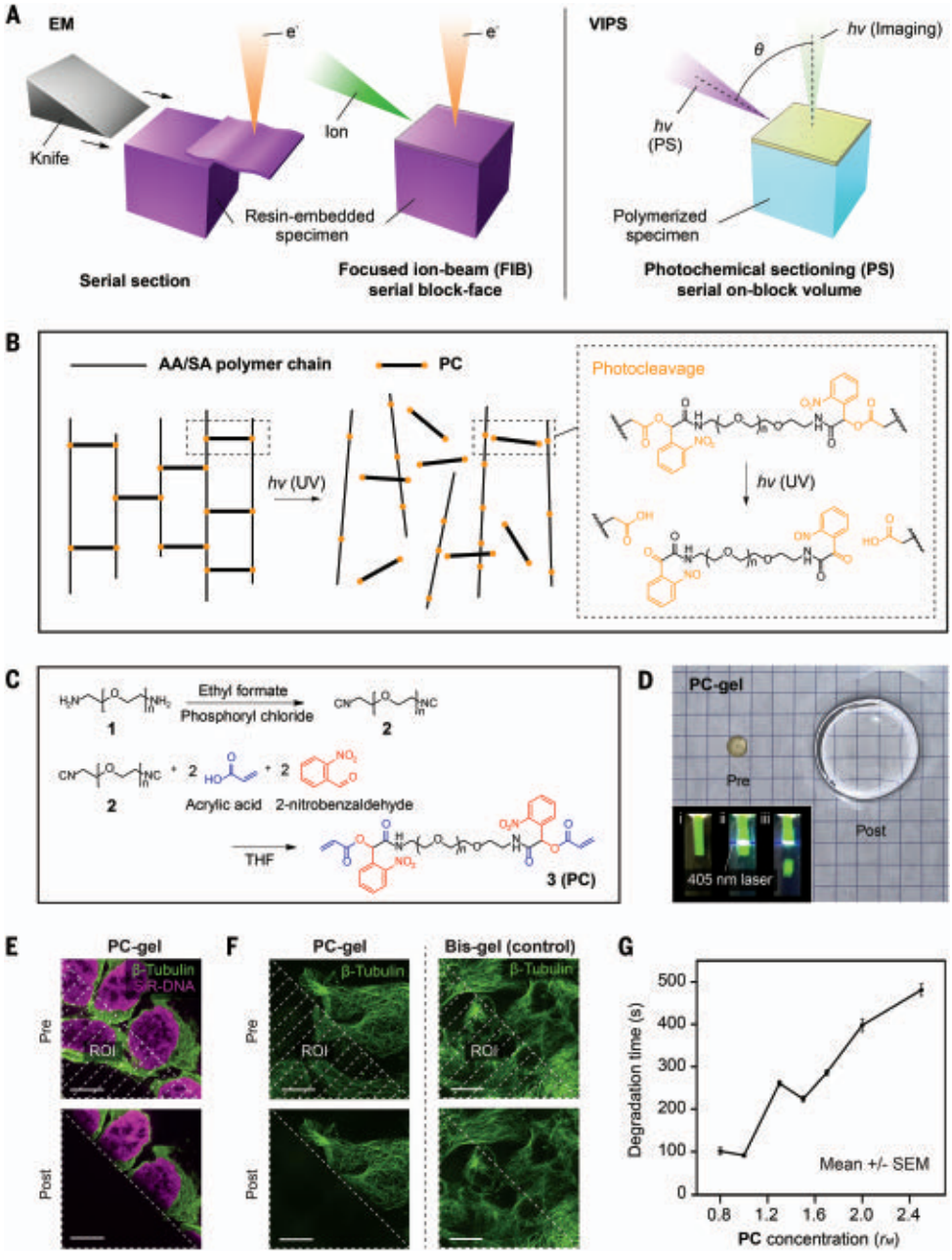


Fig. 1. VIPS. (**A**) Volumetric imaging through serial sectioning or FIB as practiced in EM (left), and volumetric optical imaging through PS in VIPS (right). The sectioning angle (θ) of PS illumination [*hv* (PS)] can be tangent or normal to the block. (**B**) Schematic of degradation of acrylamide (AA)/sodium acrylate (SA) polymer chains through photocleavage of the cross-linker (PC). Inset, detail showing photocleavage of PC under UV illumination [*hv* (UV)]. (**C**) Synthesis of PC (compound **3**). The second step uses a Passerini reaction for facile synthesis. (**D**) Pre- and post-expansion images of PC-gel. Grid size is 5 mm. Inset, PC-gel (labeled with green dye) in water before (i), during (ii), and after (iii) 405-nm laser light-sheet illumination from the side. (**E**) Pre- (top) and post- (bottom) 405-nm laser illumination images of human embryonic kidney (HEK) cells fluorescently labeled for microtubules (β-tubulin, AF488) and nuclei (SiR-DNA) and expanded with the PC-gel. The illuminated ROI at lower left was completely removed. Scale bars, 10 μm (48 μm). Hereafter, all scale bars are at pre-expansion scale, with post-expansion sizes indicated in brackets, unless otherwise noted. (**F**) Similar experiment for expanded HEK cells with β-tubulin labeling showing removal of the illuminated ROI when expanded with the PC-gel (left column) versus the nonphotodegradable Bis-gel (control) (right column). Scale bars, 10 μm (48 μm for PC-gel and 45 μm for Bis-gel). (**G**) Photodegradation time of PC-gel–expanded HEK cells under 405-nm illumination versus PC concentration ($n = 3$ samples at each concentration, r_M = molar ratio of PC to bis-acrylamide used for Bis-gel).

synthesized a photocleavable version of polyacrylamide–sodium polyacrylate hydrogel through free-radical chain-growth polymerization (Fig. 1D). The resulting optically transparent and mechanically elastic hydrogel ("PC-gel") underwent the same osmotic swelling as its nonphotodegradable counterpart ("Bis-gel," a polyacrylamide–sodium polyacrylate hydrogel cross-linked with bis-acrylamide) (*5*). Consistent with the Bis-gel, the expansion factor of the PC-gel increased as the PC concentration decreased during polymerization (fig. S5) (*5*). In practice, ~4- to 5-fold expanded samples were the most convenient to handle because the PC-gel maintained a high mechanical integrity around these expansion factors. Unless otherwise noted, we used close to this optimal expansion factor for all biological sample preparations. The global

expansion errors of the PC-gel were within ~0.3 to 1.3% over 20 μm, comparable to the Bis-gel (fig. S6) (*5*).

Spatially precise photodegradation of a PC-gel

Under continuous UV (365-nm) illumination, the bulk PC-gel underwent complete photodegradation, transforming into a disintegrated liquid state (fig. S7). Although PC photocleavage occurs most efficiently around 270 to 350 nm (*29*), the more commonly used 405-nm light was also effective, requiring ~80 min at ~0.01 W/cm^2 to cleave the gel in water using a ~200-μm-thick light sheet (Fig. 1D, inset, and Movie 2). We could also photodegrade PC-gel–expanded cell samples across a defined region of interest (ROI) with a confocal microscope (Fig. 1E and table S1). The Bis-gel controls confirmed that the signal loss within the photodegraded ROI resulted from the diffusion of freed, originally gel-anchored fluorophores rather than photobleaching (Fig. 1F and fig. S8). Finally, gelled samples with higher cross-linker (PC) density required longer illumination times at the same laser power to complete the photodegradation (Fig. 1G).

Volumetric fluorescence imaging through two-photon PS

Single-photon wide-field or confocal illumination is typically not suitable for PS in a single-objective microscope setup because out-of-focus light passes through the PC-gel and causes undesirable photodegradation outside of the intended plane. One option is to exploit the nonlinear optical properties of two-photon (2P) absorption, which photodegrades only those regions through which a raster-scanned 2P-focused spot travels. A large volume can thus be imaged with 2P VIPS by repeated rounds of: (i) imaging the sample to a depth less than the working distance of the objective; (ii) photodegrading or photoslicing with 2P illumination to remove the sample to a depth slightly shallower than that of the imaging (to leave an overlap region for image stitching); and (iii) iterating this process until the desired sample volume is imaged (Fig. 2A).

To demonstrate this, we first used a 740-nm 2P laser to photodegrade a 3D ROI within a densely labeled PC-gel–expanded mouse tissue sample and then used single-photon confocal microscopy to image the encompassing volumes after photodegradation (Fig. 2B) (*29*). The lateral and axial photodegradation boundaries matched well with the defined 3D ROI (Fig. 2, C and D). The PC-gel remained unaffected at 920-nm excitation (fig. S9), demonstrating that traditional 2P imaging of visible-range fluorophores is compatible with 2P VIPS. Next, we applied 2P VIPS to a PC-gel–expanded mouse brain strip (Fig. 2E). We performed three rounds of on-block volumetric confocal microscopy imaging covering depths of ~1.5, ~1.5, and ~1.8 mm, respectively, with each imaging round separated by ~1.0 mm of 2P photodegradation. The three stitched subvolumes spanned an ~3.8-mm-thick tissue volume, well beyond the objective's working distance of ~2.6 mm (Fig. 2, F and G).

Volumetric fluorescence imaging through light-sheet PS

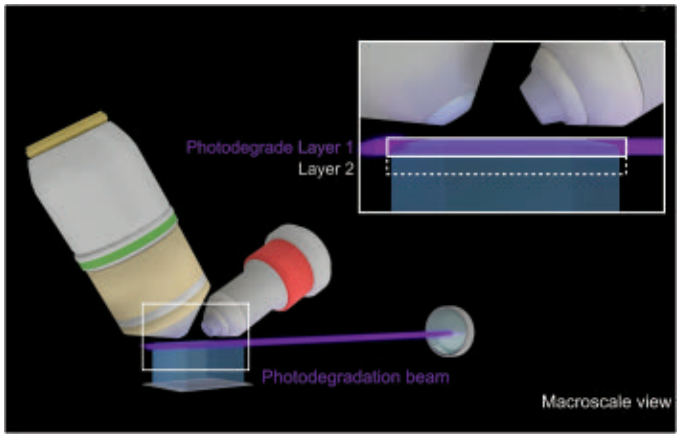
As point-scanning methods, volumetric confocal microscopy imaging and 2P PS are too slow and confocal's propensity for photobleaching is too great for either of them to be easily applied to centimeter-scale volumetric imaging. Light-sheet microscopy negates these issues by confinement of fluorescence excitation to the focal plane and parallel readout of all pixels within this plane. ExLLSM in particular has the requisite speed and low photobleaching for mesoscale imaging at nanoscale resolution (*9*, *33*), although this is limited by the effective working distances of its two objectives. It nevertheless makes an ideal platform for VIPS (Fig. 3A, fig. S10, and Movie 1).

To achieve spatially precise PS parallel to the sample surface, we integrated a separate Gaussian light-sheet illumination path (fig. S11). An initial benchmark using an ~80-mW, 405-nm laser (spread over a uniform ~70- to 100-μm-thick light sheet across an ~20 mm by 20 mm plane, ~0.02 W/cm^2) showed that photoslicing of a single ~6 mm by 10 mm PC-gel–embedded sample blockface could be completed in ~20 min for ~4- to 5-fold–expanded samples in water, and in ~2 hours for ~2-fold–expanded samples in 1× phosphate-buffered saline (PBS) as a result of the increased cross-linker density (see the materials and methods). Further characterization of sample features in the photosliced plane showed uniform surface flatness, minor distortion on par with the baseline expansion errors, and negligible photobleaching from the photoslicing process itself (figs. S12 to S14).

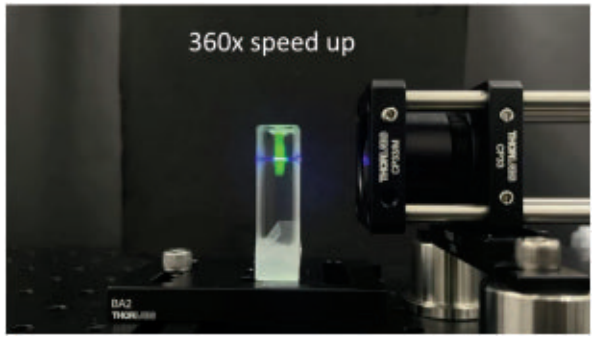
We then validated the capabilities of LLSM VIPS on an ~4.4-fold–expanded human hippocampus tissue sample labeled for axons and myelin sheaths. We performed eight rounds of sequential volumetric LLSM imaging followed by photoslicing across an ~0.25 mm by 0.23 mm by 1.6-mm volume, achieving five times the depth attainable without VIPS (fig. S15). Imaging and photoslicing times totaled ~15 min and ~2 hours, respectively, with each photosliced top gel block floated off and removed from the imaging bath immediately. The imaged subvolumes were then deconvolved, registered, and stitched using PetaKit5D with an effective Nyquist-sampled voxel size of ~22 nm by 22 nm by 59 nm (*34*). In the stitched volume, we were able to trace and reconstruct individual axons with diameters as small as ~150 nm, confirming the preservation of connectivity and traceability of anatomical features across multiple rounds of imaging and PS and providing the motivation to apply this to even larger volumes.

Mapping axons and myelin sheaths in whole mouse OB

The mammalian OB is a complex neural structure essential for processing and relaying olfactory information. To demonstrate the scalability of LLSM VIPS, we performed two-color imaging of an entire 7-week-old



Movie 1. Animation showing the concept and workflow of VIPS in both macroscale and nanoscale views.



Movie 2. Accelerated video showing spatially controlled PS of a PC-gel with a thin sheet of 405-nm laser from the side.

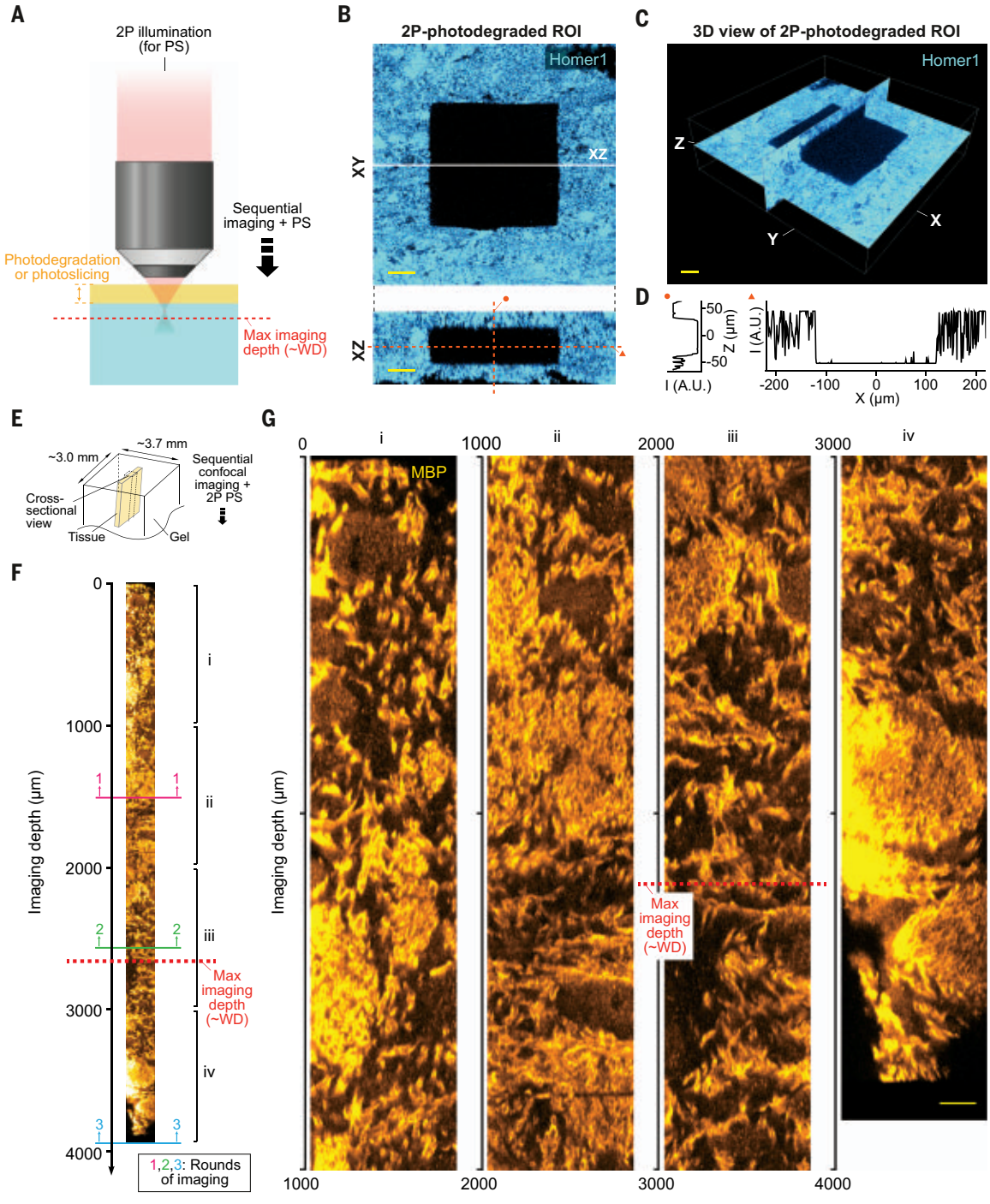


Fig. 2. Volumetric fluorescence imaging of mouse brain tissue at nanoscale resolution with 2P VIPS. (**A**) Schematic of the optical setup for 2P VIPS. Imaging beyond the microscope working distance (WD) is achieved by sequential volumetric fluorescence imaging and 2P PS. (**B**) XY (top) and XZ (bottom) views of a 2P-photodegraded ROI in mouse brain tissue fluorescently labeled for synaptic proteins (Homer1) and expanded with the PC-gel. Scale bars, 50 μm (post-expansion). (**C**) 3D-rendered view of the 2P-photodegraded ROI in (B). (**D**) Profiles of fluorescence intensity along the orange dotted lines in (B). (**E**) Schematic of a mouse brain tissue strip vertically embedded and expanded with the PC-gel for imaging by 2P VIPS. (**F**) Vertical orthoslice through the tissue strip imaged to an ~3.8-mm depth (beyond the ~2.6-mm objective WD) through three rounds of sequential on-block volumetric confocal microscopy imaging and 2P PS. Myelin sheaths, labeled with fluorescent MBP, are shown. Lines 1 to 3 indicate the maximum imaging depth for each round. (**G**) Enlarged views of regions i to iv as denoted in (F). Scale bar, 50 μm (post-expansion).

WT mouse OB at ~2-fold expansion (~56 nm by 56 nm by 135 nm effective voxel size). Although modest, this expansion factor yielded resolution sufficient to trace and reconstruct most axons in the mouse brain, as estimated from volume EM (*35*), and avoids the tens of petabytes of data for the raw, intermediate, and fully processed volumes that would be necessary at higher expansion factors.

We imaged the entire expanded OB (~2-fold expanded to a volume of ~6.4 mm by 9.4 mm by 4.3 mm) at peak acquisition rates of ~2 teravoxels per hour over 10 days using nine rounds of interleaved on-block volumetric LLSM imaging and photoslicing (Fig. 3B). Imaging and photoslicing times totaled ~142 and ~20 hours, respectively, and yielded 17,280 volumetric tiles, constituting ~0.5 petabytes of raw data.

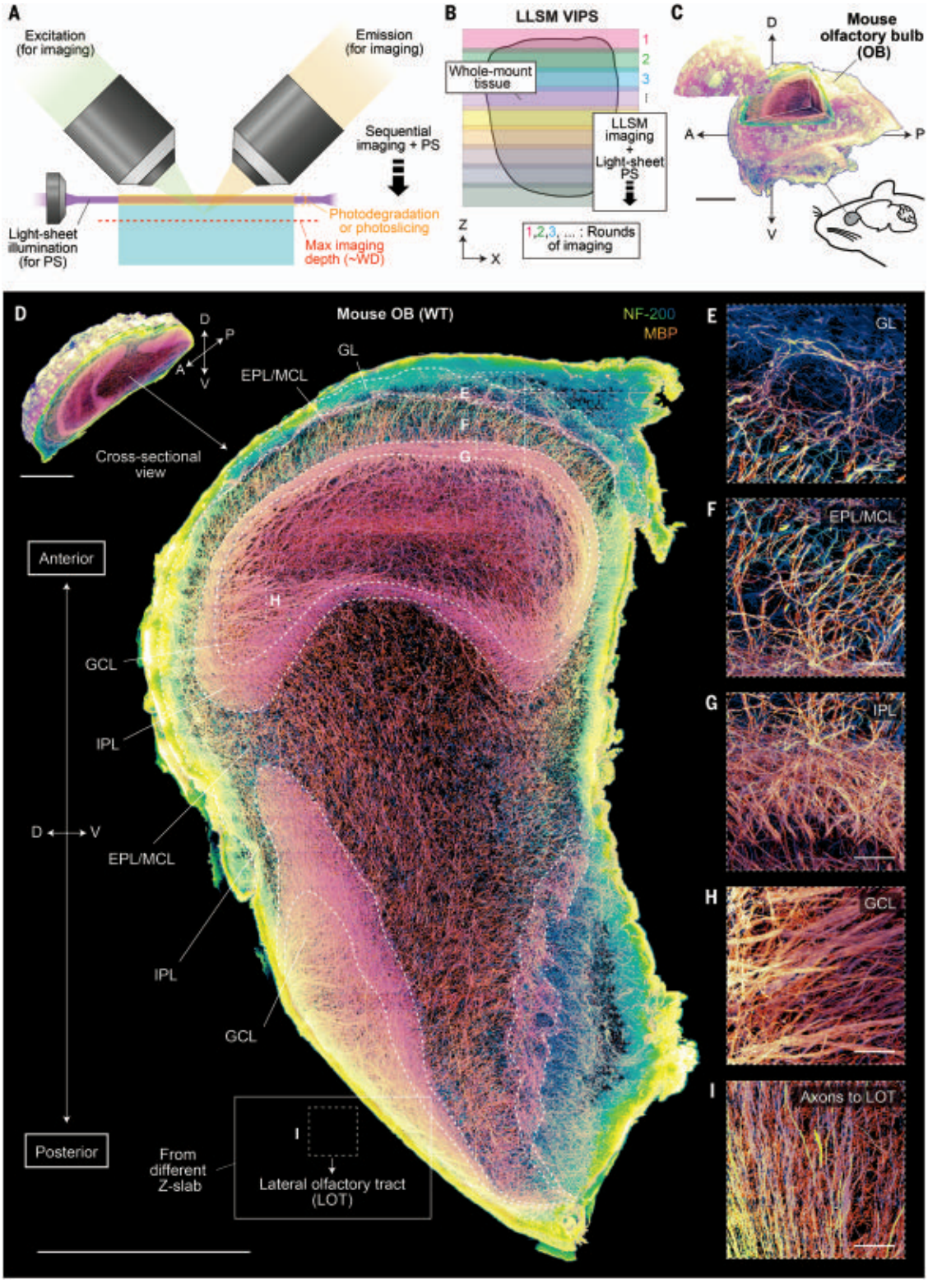


Fig. 3. Volumetric fluorescence imaging of mouse OB at nanoscale resolution with LLSM VIPS. (**A**) Schematic of the optical setup. Imaging beyond the WD of the two angled LLSM objectives was achieved by light-sheet PS with a third objective tangent to the block. (**B**) Strategy for whole OB imaging showing multiple rounds of overlapping LLSM volumetric imaging and light-sheet PS to ensure lossless reconstruction of the entire imaged volume. (**C**) Top, 3D-rendered view of a 7-week-old WT mouse OB fluorescently labeled for axons [neurofilament-200 (NF-200)] and myelin sheaths (MBP) (~3.3 mm by 4.9 mm by 2.2 mm, pre-expansion, ~6.4 mm by 9.4 mm by 4.3 mm, post-expansion, wedge removed to show the interior). Bottom, approximate location of the OB in the mouse brain. Scale bar, 1 mm (1.93 mm). A, anterior; P, posterior; D, dorsal; V, ventral (**D**) 3D-rendered slab view (~80-μm thick, pre-expansion) of the WT mouse OB, downsampled 5-fold in each direction for visualization. Scale bar, 1 mm (1.93 mm). Top inset, 3D view indicating the location in the OB of the imaged slab. Scale bar, 1 mm (1.93 mm). (**E** to **I**) Full-resolution maximum intensity projection (MIP) views (~80-μm thick, pre-expansion) of the boxed regions in (D) near: GL (E); EPL/MCL (F); IPL (G); GCL (H); and LOT (I). Region I is from a different slab view than the one shown in (D). Scale bars, 50 μm (96.5 μm).

www.xsava.xyz

We reconstructed the dataset in two stages. First, parallelized PetaKit5D processing handled de-skewing, rotation, deconvolution, and stitching to rebuild the full OB (Fig. 3C and Movie 3) (*34*). Second, PetaVIPS, an extension of PetaKit5D, enabled scalable and parallel downstream workflows, from axon and myelin segmentation to tractography analysis (see the materials and methods) (*36*).

The reconstructed WT mouse OB dataset revealed distinct axonal projections and myelination patterns across all of its major anatomical layers (*37*) (Fig. 3D). For example, detailed 3D visualization highlighted the unique axonal and myelination patterns within individual glomeruli of the glomerular layer (GL), where the axonal shafts formed a dense, peripheral basket encompassing the central neuropil (Fig. 3E). The vertical axonal projections of tufted cells traversed through the external plexiform layer (EPL) and mitral cell layer (MCL) (Fig. 3F), and the dense lateral axonal projections spanned the internal plexiform layer (IPL) (Fig. 3G). The axonal bundles from the mitral and tufted cells were observed passing through the granule cell layer (GCL) (Fig. 3H), exiting the OB, and projecting through the lateral olfactory tract (LOT) to the olfactory cortex (Fig. 3I). Overall, axons traveling through the GL and EPL/MCL exhibited high tortuosity, whereas those in the deeper layers displayed strong directionality. Furthermore, the nanoscale resolution of the dataset enabled segmentation, skeletonization, and reconstruction of individual myelinated and unmyelinated axon fibers, allowing for precise measurements such as calculating the total lengths of segmentable axons throughout the entirety of mouse OB (Fig. 4, A and B, and figs. S16 and S17).

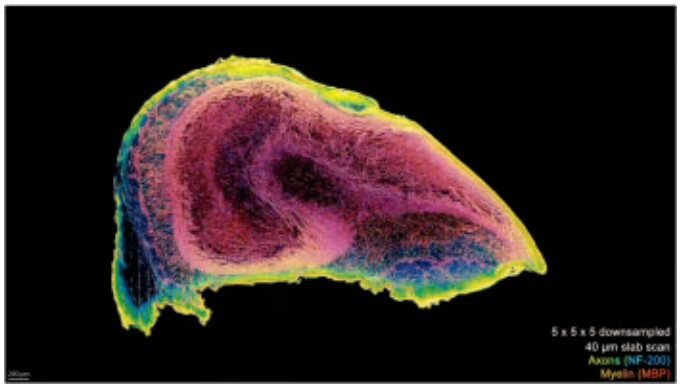
Axon degeneration and de-/dysmyelination in a neurodegenerative mouse OB

Axonal degeneration and de-/dysmyelination, which often precede clinical symptoms, are hallmarks of neurodegeneration and aging. In the fatal neurodegenerative lysosomal storage disorder Niemann-Pick type C1 (NPC1) disease, mutations in the *NPC1* gene cause early-stage dysmyelination (*38*). Pathophysiological changes in the NPC1 animal's olfactory system represent an early indicator of disease progression (*39*, *40*), which is also characterized by reduced myelin basic protein (MBP) expression in the OB at the neonatal and adolescent stage (*41*). However, the precise mechanism behind de-/dysmyelination remains unclear, particularly regarding whether the reduced expression of MBP in the OB persists into adulthood and results in anatomical loss at the single-axon or myelin sheath level. Furthermore, the progression of axon and myelin sheath loss across the distinct layers of the OB remains unexplored.

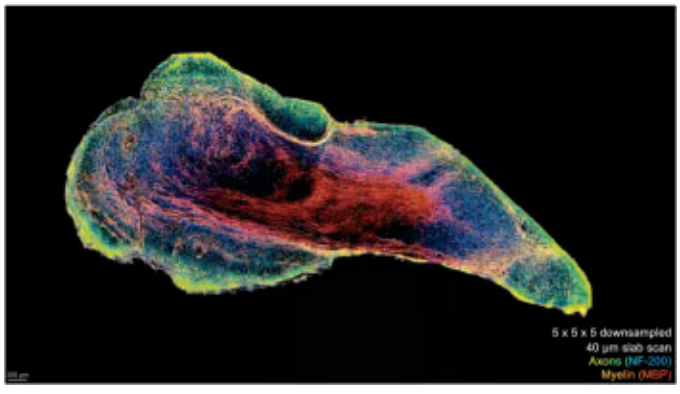
To study axon degeneration and de-/dysmyelination in the neurodegenerative brain, we imaged a 7-week-old NPC1 (*Npc1*$^{-/-}$) mouse OB (~2-fold expanded to a volume of ~6.0 mm by 11.2 mm by 3.6 mm; ~0.6 petabytes of raw data) (Movie 4). The expanded and reconstructed WT and NPC1 mouse OB volumes enabled direct comparison of axon and myelin sheath morphologies across the entire OB (Movie 3 and 4 and Fig. 4C), which revealed distinct anatomical changes in the NPC1 OB. These included apparent reduction in axon and myelin sheath densities in the GL, EPL/MCL, and IPL, indicative of potential axon degeneration and de-/dysmyelination in these OB layers (Fig. 4C).

An OB-wide quantitative analysis of axon fibers revealed a reduced volume density of segmented axons in the NPC1 OB, consistent with previous studies showing early-onset axonal degeneration across all major parts of the NPC1 brain (Fig. 4D) (*42*–*44*). The severity of axon degeneration varied across different OB layers, with the EPL/MCL exhibiting the greatest reduction (–56.6%) and the IPL showing the least (–29.6%). An exception was in the GCL, which exhibited an increase in axon density (+18.6%) (Fig. 4D). We hypothesize that this may reflect compensatory axonal sprouting in the GCL after widespread neuronal loss (*45*). Given that axonal damage often precedes loss of cell bodies (*43*, *46*), which can be located in distant anatomical layers and brain regions, these results highlight VIPS's capability to capture the local variability of axonal integrity and survival both within and across the OB layers.

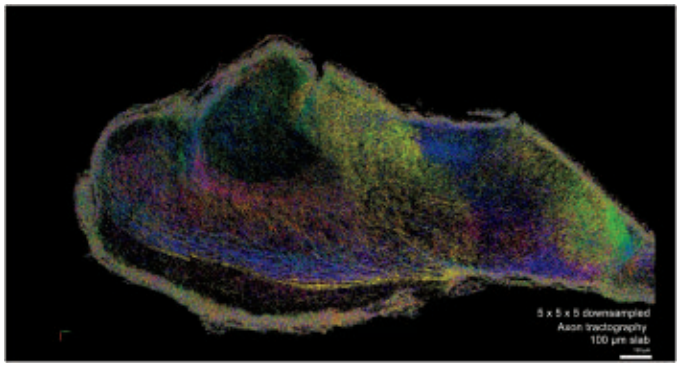
We also compared myelination profiles of the WT and NPC1 mouse OBs. The density of total segmented myelin sheaths and myelinated axons was reduced throughout the NPC1 OB (Fig. 4E), affirming that early-stage dysmyelination persists as the animal develops (*41*). Differences in the percentage of myelinated axons to all axons also varied across the layers. Only marginal or no decreases from the WT to the NPC1 OB were observed in the GL and the EPL/MCL (Fig. 4F), indicating a similar rate of loss for both myelinated and unmyelinated axons in these layers. By contrast, the IPL and GCL showed decreases of 36.2 and 46.9%, respectively, suggesting that myelinated axons in these layers may have experienced greater loss than unmyelinated ones (*47*). Such spatial heterogeneity was also pronounced within each layer,



Movie 3. Volumetric rendering showing a 7-week-old WT mouse OB imaged with LLSM VIPS.



Movie 4. Volumetric rendering showing a 7-week-old NPC1 (*Npc1*$^{-/-}$) mouse OB imaged with LLSM VIPS.



Movie 5. Volumetric rendering showing the ϕ angles of segmented axons in the 7-week-old WT and NPC1 mouse OBs.

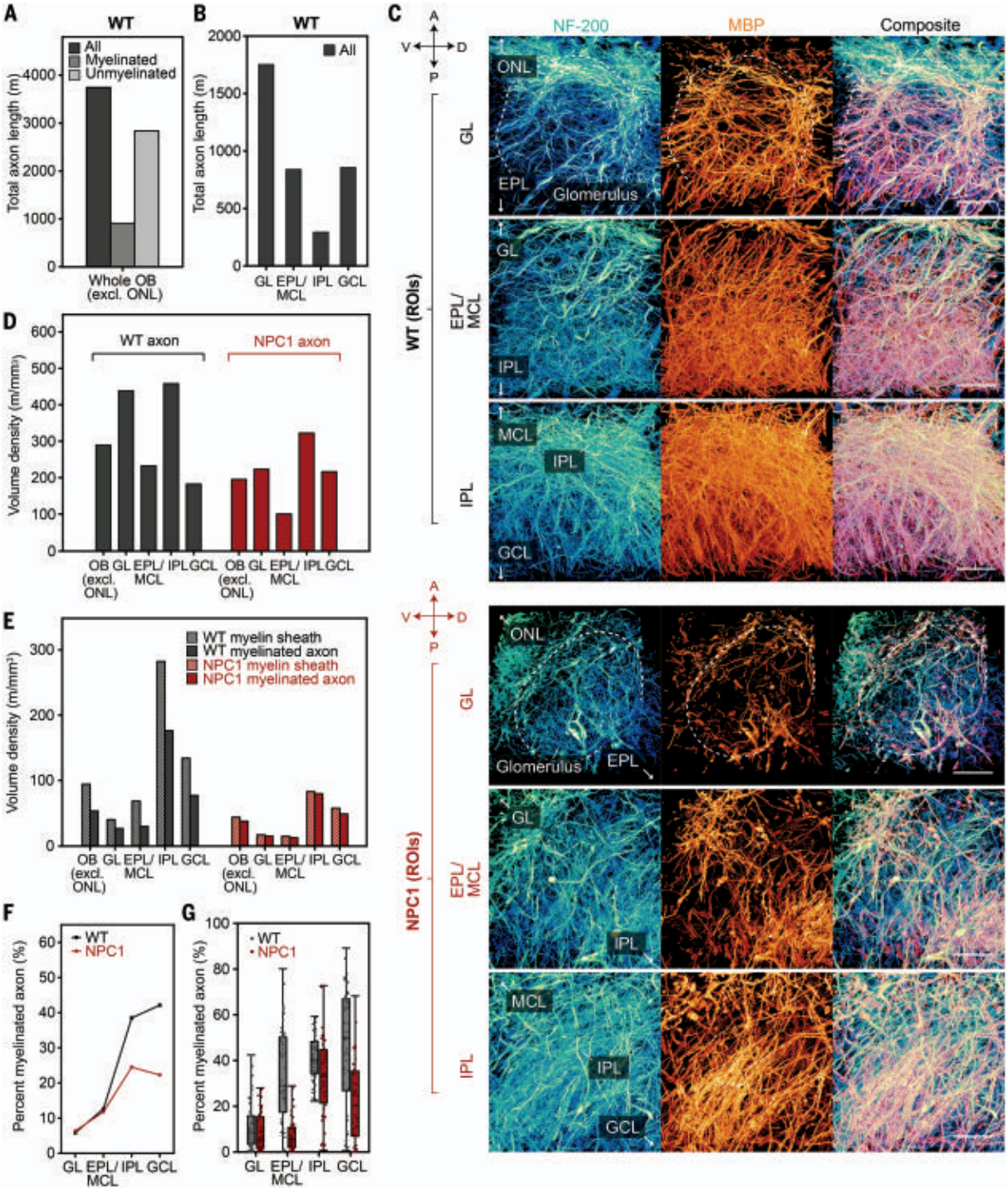


Fig. 4. OB-wide analysis of axons and myelin sheaths at single-axon resolution. (**A**) Total length of segmented axons, myelinated axons, and unmyelinated axons across a WT mouse OB [excluding the olfactory nerve layer (ONL)]. (**B**) Total length of segmented axons in distinct anatomical layers. (**C**) Full-resolution 3D-rendered views of representative ROIs in the GL, EPL/MCL, and IPL of 7-week-old WT (top) and NPC1 (*Npc1*$^{-/-}$) (bottom) mouse OBs. NF-200, MBP, and composite views of the same region are shown in separate columns. Scale bars, 50 μm (96.5 μm). (**D**) Comparison of the volume density of segmented axons across the entirety of and in distinct anatomical layers of the WT and NPC1 mouse OBs. (**E**) Similar comparison of segmented myelin sheaths and myelinated axons across the two OBs. (**F**) Percentage of myelinated axons relative to all segmented axons across distinct anatomical layers of the two OBs. (**G**) Box plots of the percentage of myelinated axons across ~50 ROIs (~66 μm by 66 μm by 87 μm, pre-expansion) evenly sampled across distinct anatomical layers of the WT and NPC1 mouse OBs. Whisker ends denote maximum and minimum values; dots denote individual values; and the upper, middle, box lines denote the 75th, 50th (median), and 25th percentiles, respectively.

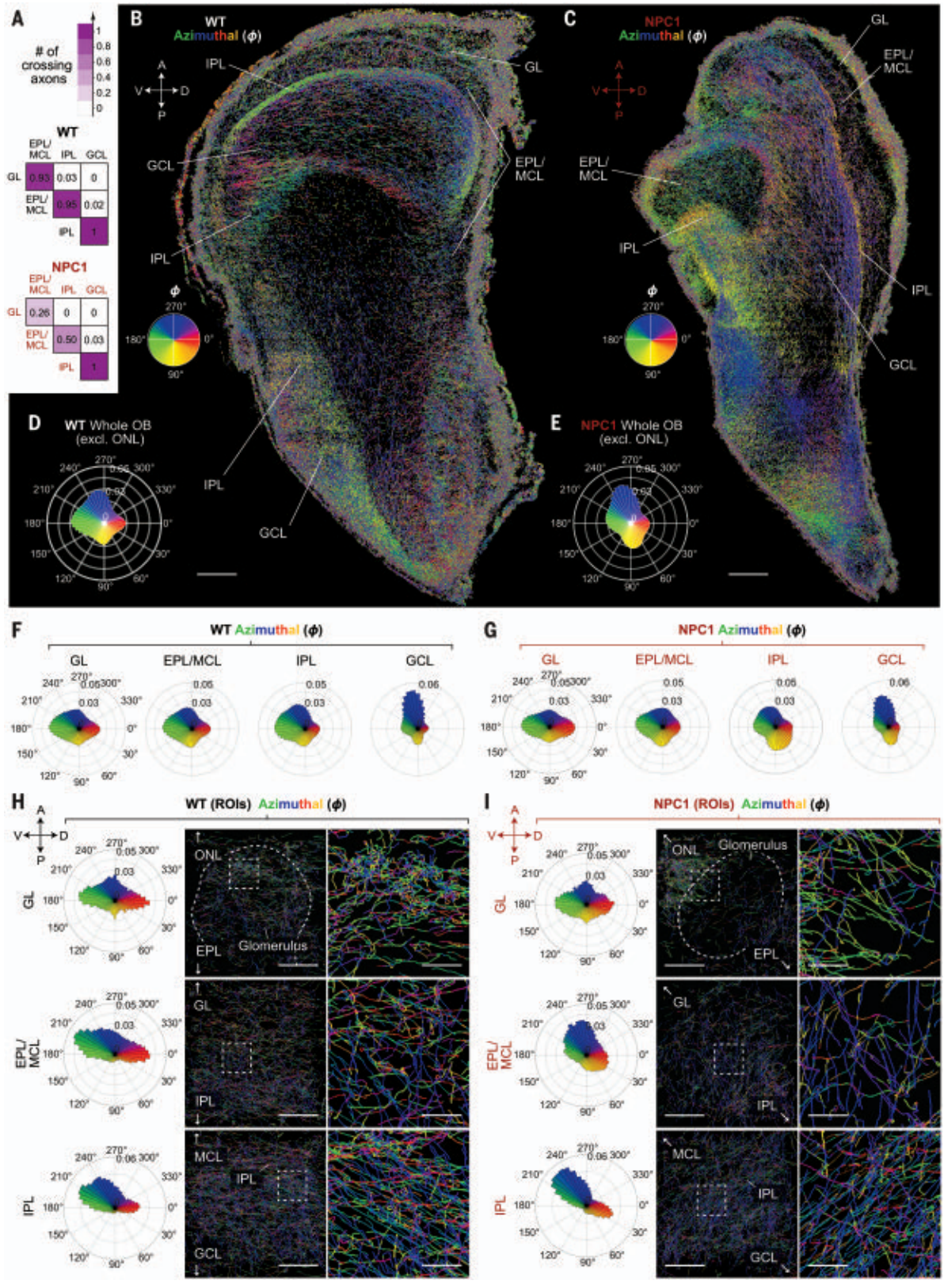


Fig. 5. OB-wide topography and tractography at single-axon resolution. (**A**) Total number of segmented axons crossing distinct anatomical layers in 7-week-old WT (left) and NPC1 (*Npc1*$^{-/-}$) (right) mouse OBs. Axon numbers are normalized to those crossing between the IPL and GCL in each OB. (**B** and **C**) 3D-rendered slab views (~80-μm thick, pre-expansion) of azimuthal (ϕ) angles of segmented axons in the two OBs downsampled 10-fold in each direction for visualization. Scale bars, 200 μm (386 μm). (**D** and **E**) Frequency histograms of ϕ angles of all axons segmented across the entirety of the two OBs (excluding the ONL). (**F** and **G**) Frequency histograms of ϕ angles of all axons segmented across distinct anatomical layers of the two OBs. (**H** and **I**) Frequency histograms and full-resolution 3D-rendered views of ϕ angles of segmented axons in representative ROIs in GL, EPL/MCL, and IPL of the two OBs. Scale bars, 50 μm (96.5 μm) and 10 μm (19.3 μm) for the enlarged boxed regions.

with the myelination characteristics varied widely across the spatially sampled ROIs (Fig. 4G and fig. S18).

OB-wide topography and tractography

Leveraging the nanoscale resolution and whole-volume continuity provided by VIPS, we performed single-axon topography and tractography to evaluate axonal connectivity across the WT and NPC1 OB volumes. In the WT OB, the number of segmented axons traversing between adjacent layers remained constant throughout the bulb (Fig. 5A). In the NPC1 OB, however, progressively fewer segmented axons traveled from the IPL-GCL to the EPL/MCL-IPL and then to the GL-EPL/MCL interfaces, indicating that degeneration begins in the outer layers and advances inward.

Single-axon tractography enabled visualization (Fig. 5, B and C; figs. S19 and S20; and Movie 5) and quantitative analysis (Fig. 5, D and E, and fig. S21) of the 3D organization of the segmented axons across the entire OB, as well as within specific OB layers (Fig. 5, F and G, and fig. S21) and localized regions (Fig. 5, H and I, and fig. S22). Within each layer, axon orientations shifted from a ventral-dorsal axis in the outer layer (e.g., GL) to an anterior-posterior axis in the deeper layer (e.g., GCL), reflecting the shifting paths of information flow (Fig. 5, F to I, and fig. S21). Compared with the WT OB, the NPC1 OB showed the most prominent reduction in axons projecting along the ventral-dorsal axis within the IPL, suggesting that the degenerative process can be selective for axons with specific orientations. Within this overall architecture, however, there existed substantial local heterogeneity, with axons projecting in many directions even if the average orientations in each region followed an overall trend (Fig. 5, F to I).

Finally, in the WT OB, the myelinated axon orientations mirrored those of the total axon population, whereas in the IPL and GCL of the NPC1 OB, the myelinated axons were strongly biased along the anterior-posterior axis (Fig. 5, F and G, and fig. S21). This suggests that the remaining axons in these layers, mostly myelinated, project primarily to or from the olfactory cortex through the LOT (*48*).

Discussion

VIPS overcomes the working distance limits of high-resolution optical microscopy by repeated rounds of in situ PS and imaging of a hydrogel-embedded sample. This noncontact process enables precise sectioning compatible with various hydrogel-expansion chemistries and circumvents the distortion, sample loss, and protocol limitations of mechanical sectioning (see supplementary note 1). Furthermore, the process can be easily integrated into existing microscope setups, enabling continuous automated acquisition. When combined with LLSM imaging (*33*) and a high-performance computational pipeline (*34*, *36*), VIPS effectively eliminates physical sample size as a barrier to mesoscale imaging of tissues at nanoscale resolution.

The large volumetric dynamic range of VIPS reveals anatomical heterogeneities, gradients, and biological insights that are undetectable when sampling smaller volumes lacking global context or spatial resolution (see supplementary note 2). Although only a single WT and NPC1 pair of OBs were compared in our study, a dedicated pipeline based on the sample preparation, imaging, and computation methods demonstrated here has the potential bandwidth to image multiple specimens to quantitatively address the degree of stereotypy and to explore different molecular markers to achieve more biologically meaningful comparisons and overall understanding.

Extending VIPS to the petabyte scale has exposed a series of technical challenges (see supplementary notes 3 and 4). First, labeling density and specificity proved limiting even at only ~2-fold expansion, where our segmented count of thin unmyelinated axons in the LOT was only ~14% of that tallied in a prior EM study of the same pathway, albeit at a different animal age (*49*). Dense and uniform labeling may prove even more challenging in larger specimens. Second, the nonlinear warping observed at subvolume boundaries, which might be expected to increase at higher expansion factors, exposes the need for scalable, nonrigid registration pipelines capable of handling peta- to exavoxel data. Third, even with real-time compression, single samples may exceed a petabyte; storage costs, data processing, analysis, and manual curation may set practical limits on the number and size of specimens long before the limits of imaging are reached. Finally, 3D petabyte-scale data are far too vast to be mined effectively and comprehensively for new insights by humans alone. AI systems will be needed to replace manual annotation and to extract quantitative information at scale under human queries. Progress in labeling chemistry, AI/GPU–accelerated nonrigid solvers and segmentation, compute costs, and AI vision are already narrowing these gaps.

Despite these challenges, by greatly extending the tissue volumes approachable with nanoscale fluorescence microscopy, VIPS takes a major step in advancing from studies that attempt to derive general insights from sparse subsampling of large tissues to quantitative measurement of biological stereotypy and variability and the derivation of fundamental insights from tissues in toto. Indeed, coupled with pan-membrane or protein labeling, VIPS might form the foundation for an optical pipeline to deliver synaptic-level dense connectomes of whole brains (*50*, *51*).

Materials and methods are available as supplementary materials.

REFERENCES AND NOTES

1. C. Kirst *et al.*, Mapping the fine-scale organization and plasticity of the brain vasculature. *Cell* **180**, 780–795.e25 (2020). doi: 10.1016/j.cell.2020.01.028; pmid: 32059781
2. C. Zhou *et al.*, Continuous subcellular resolution three-dimensional imaging on intact macaque brain. *Sci. Bull.* **67**, 85–96 (2022). doi: 10.1016/j.scib.2021.08.003; pmid: 36545964
3. D. S. Richardson, J. W. Lichtman, Clarifying Tissue Clearing. *Cell* **162**, 246–257 (2015). doi: 10.1016/j.cell.2015.06.067; pmid: 26186186
4. H. R. Ueda *et al.*, Tissue clearing and its applications in neuroscience. *Nat. Rev. Neurosci.* **21**, 61–79 (2020). doi: 10.1038/s41583-019-0250-1; pmid: 31896771
5. F. Chen, P. W. Tillberg, E. S. Boyden, Expansion microscopy. *Science* **347**, 543–548 (2015). doi: 10.1126/science.1260088; pmid: 25592419
6. W. Wang, Y. H. Chan, S. Kwon, J. Tandukar, R. Gao, Nanoscale fluorescence imaging of biological ultrastructure via molecular anchoring and physical expansion. *Nano Converg.* **9**, 30 (2022). doi: 10.1186/s40580-022-00318-6; pmid: 35810234
7. V. Gradinaru, J. Treweek, K. Overton, K. Deisseroth, Hydrogel-tissue chemistry: Principles and applications. *Annu. Rev. Biophys.* **47**, 355–376 (2018). doi: 10.1146/annurev-biophys-070317-032905; pmid: 29792820
8. S. W. Choi, W. Guan, K. Chung, Basic principles of hydrogel-based tissue transformation technologies and their applications. *Cell* **184**, 4115–4136 (2021). doi: 10.1016/j.cell.2021.07.009; pmid: 34358468
9. R. Gao *et al.*, Cortical column and whole-brain imaging with molecular contrast and nanoscale resolution. *Science* **363**, eaau8302 (2019). doi: 10.1126/science.aau8302; pmid: 30655415
10. S. Daetwyler, R. P. Fiolka, Light-sheets and smart microscopy, an exciting future is dawning. *Commun. Biol.* **6**, 502 (2023). doi: 10.1038/s42003-023-04857-4; pmid: 37161000
11. J. L. Lillvis *et al.*, Rapid reconstruction of neural circuits using tissue expansion and light sheet microscopy. *eLife* **11**, e81248 (2022). doi: 10.7554/eLife.81248; pmid: 36286237
12. X. Tian *et al.*, Rapid lightsheet fluorescence imaging of whole Drosophila brains at nanoscale resolution by potassium acrylate-based expansion microscopy. *Nat. Commun.* **15**, 10911 (2024). doi: 10.1038/s41467-024-55305-8; pmid: 39738207
13. Y. Zhang, H. Gross, Systematic design of microscope objectives. Part I: System review and analysis. *Adv. Opt. Technol.* **8**, 313–347 (2019). doi: 10.1515/aot-2019-0002
14. A. Glaser *et al.*, Expansion-assisted selective plane illumination microscopy for nanoscale imaging of centimeter-scale tissues. *eLife* **12**, RP91979 (2025). doi: 10.7554/eLife.91979; pmid: 40586775
15. W. Yin *et al.*, A petascale automated imaging pipeline for mapping neuronal circuits with high-throughput transmission electron microscopy. *Nat. Commun.* **11**, 4949 (2020). doi: 10.1038/s41467-020-18659-3; pmid: 33009388
16. L. K. Scheffer *et al.*, A connectome and analysis of the adult *Drosophila* central brain. *eLife* **9**, e57443 (2020). doi: 10.7554/eLife.57443; pmid: 32880371
17. C. J. Peddie *et al.*, Volume electron microscopy. *Nat. Rev. Methods Primers* **2**, 51 (2022). doi: 10.1038/s43586-022-00131-9; pmid: 37409324
18. A. Shapson-Coe *et al.*, A petavoxel fragment of human cerebral cortex reconstructed at nanoscale resolution. *Science* **384**, eadk4858 (2024). doi: 10.1126/science.adk4858; pmid: 38723085

19. H. Gong *et al.*, High-throughput dual-colour precision imaging for brain-wide connectome with cytoarchitectonic landmarks at the cellular level. *Nat. Commun.* **7**, 12142 (2016). doi: 10.1038/ncomms12142; pmid: 27374071
20. J. Winnubst *et al.*, Reconstruction of 1,000 projection neurons reveals new cell types and organization of long-range connectivity in the mouse brain. *Cell* **179**, 268–281.e13 (2019). doi: 10.1016/j.cell.2019.07.042; pmid: 31495573
21. W. Yu *et al.*, Translational rapid ultraviolet-excited sectioning tomography for whole-organ multicolor imaging with real-time molecular staining. *eLife* **11**, e81015 (2022). doi: 10.7554/eLife.81015; pmid: 36331195
22. Y. Yi *et al.*, Mapping of individual sensory nerve axons from digits to spinal cord with the transparent embedding solvent system. *Cell Res.* **34**, 124–139 (2024). doi: 10.1038/s41422-023-00867-3; pmid: 38168640
23. F. Xu *et al.*, High-throughput mapping of a whole rhesus monkey brain at micrometer resolution. *Nat. Biotechnol.* **39**, 1521–1528 (2021). doi: 10.1038/s41587-021-00986-5; pmid: 34312500
24. P. S. Tsai *et al.*, All-optical histology using ultrashort laser pulses. *Neuron* **39**, 27–41 (2003). doi: 10.1016/S0896-6273(03)00370-2; pmid: 12848930
25. M.-Y. Shi *et al.*, High-speed mapping of whole-mouse peripheral nerves at subcellular resolution. *Cell* **188**, 3897–3915.e20 (2025). doi: 10.1016/j.cell.2025.06.011; pmid: 40645171
26. R. Chen *et al.*, Expansion tomography for large volume tissue imaging with nanoscale resolution. *Biomed. Opt. Express* **12**, 5614–5628 (2021). doi: 10.1364/BOE.431696; pmid: 34692204
27. R. Torres *et al.*, A scalable and modular computational pipeline for axonal connectomics: automated tracing and assembly of axons across serial sections. bioRxiv [Preprint] 598365 (2024); https://doi.org/10.1101/2024.06.11.598365/.
28. J. Park *et al.*, Integrated platform for multiscale molecular imaging and phenotyping of the human brain. *Science* **384**, eadh9979 (2024). doi: 10.1126/science.adh9979; pmid: 38870291
29. A. M. Kloxin, A. M. Kasko, C. N. Salinas, K. S. Anseth, Photodegradable hydrogels for dynamic tuning of physical and chemical properties. *Science* **324**, 59–63 (2009). doi: 10.1126/science.1169494; pmid: 19342581
30. F. Yanagawa *et al.*, Activated-ester-type photocleavable crosslinker for preparation of photodegradable hydrogels using a two-component mixing reaction. *Adv. Healthc. Mater.* **4**, 246–254 (2015). doi: 10.1002/adhm.201400180; pmid: 25116476
31. J. M. Scheiger *et al.*, Inherently UV photodegradable poly(methacrylate) gels. *Adv. Funct. Mater.* **31**, 2105681 (2021). doi: 10.1002/adfm.202105681
32. K. A. Günay *et al.*, Photo-expansion microscopy enables super-resolution imaging of cells embedded in 3D hydrogels. *Nat. Mater.* **22**, 777–785 (2023). doi: 10.1038/s41563-023-01558-5; pmid: 37217701
33. T.-M. Fu *et al.*, A multimodal adaptive optical microscope for in vivo imaging from molecules to organisms. bioRxiv [Preprint] 657494 (2025); https://doi.org/10.1101/2025.06.02.657494.
34. X. Ruan *et al.*, Image processing tools for petabyte-scale light sheet microscopy data. *Nat. Methods* **21**, 2342–2352 (2024). doi: 10.1038/s41592-024-02475-4; pmid: 39420143
35. G. S. Tomassy *et al.*, Distinct profiles of myelin distribution along single axons of pyramidal neurons in the neocortex. *Science* **344**, 319–324 (2014). doi: 10.1126/science.1249766; pmid: 24744380
36. X. Ruan, E. Betzig, S. Upadhyayula, "abcucberkeley/PetaVIPS: v1.0.0" (Zenodo, 2025); doi: 10.5281/zenodo.16929810
37. F. Imamura, A. Ito, B. J. LaFever, Subpopulations of projection neurons in the olfactory bulb. *Front. Neural Circuits* **14**, 561822 (2020). doi: 10.3389/fncir.2020.561822; pmid: 32982699
38. T. Yu, A. P. Lieberman, Npc1 acting in neurons and glia is essential for the formation and maintenance of CNS myelin. *PLOS Genet.* **9**, e1003462 (2013). doi: 10.1371/journal.pgen.1003462; pmid: 23593041
39. M. Hovakimyan *et al.*, Olfactory deficits in Niemann-Pick type C1 (NPC1) disease. *PLOS ONE* **8**, e82216 (2013). doi: 10.1371/journal.pone.0082216; pmid: 24391715
40. A. Meyer *et al.*, Olfactory performance as an indicator for protective treatment effects in an animal model of neurodegeneration. *Front. Integr. Neurosci.* **12**, 35 (2018). doi: 10.3389/fnint.2018.00035; pmid: 30154701
41. X. Yan *et al.*, Decreased expression of myelin gene regulatory factor in Niemann-Pick type C 1 mouse. *Metab. Brain Dis.* **26**, 299–306 (2011). doi: 10.1007/s11011-011-9263-9; pmid: 21938520
42. W.-Y. Ong *et al.*, Neurodegeneration in Niemann-Pick type C disease mice. *Exp. Brain Res.* **141**, 218–231 (2001). doi: 10.1007/s002210100870; pmid: 11713633
43. S. Takikita, T. Fukuda, I. Mohri, T. Yagi, K. Suzuki, Perturbed myelination process of premyelinating oligodendrocyte in Niemann-Pick type C mouse. *J. Neuropathol. Exp. Neurol.* **63**, 660–673 (2004). doi: 10.1093/jnen/63.6.660; pmid: 15217094
44. S. N. R. Pressey, D. A. Smith, A. M. S. Wong, F. M. Platt, J. D. Cooper, Early glial activation, synaptic changes and axonal pathology in the thalamocortical system of Niemann-Pick type C1 mice. *Neurobiol. Dis.* **45**, 1086–1100 (2012). doi: 10.1016/j.nbd.2011.12.027; pmid: 22198570
45. K. L. Marshall, M. H. Farah, Axonal regeneration and sprouting as a potential therapeutic target for nervous system disorders. *Neural Regen. Res.* **16**, 1901–1910 (2021). doi: 10.4103/1673-5374.308077; pmid: 33642358
46. S. C. Patel *et al.*, Localization of Niemann-Pick C1 protein in astrocytes: Implications for neuronal degeneration in Niemann- Pick type C disease. *Proc. Natl. Acad. Sci. USA.* **96**, 1657–1662 (1999). doi: 10.1073/pnas.96.4.1657; pmid: 9990080
47. U. Fünfschilling *et al.*, Glycolytic oligodendrocytes maintain myelin and long-term axonal integrity. *Nature* **485**, 517–521 (2012). doi: 10.1038/nature11007; pmid: 22622581
48. D. Brunert, M. Rothermel, Extrinsic neuromodulation in the rodent olfactory bulb. *Cell Tissue Res.* **383**, 507–524 (2021). doi: 10.1007/s00441-020-03365-9; pmid: 33355709
49. P. C. Brunjes, R. B. Kay, J. P. Arrivillaga, The mouse olfactory peduncle. *J. Comp. Neurol.* **519**, 2870–2886 (2011). doi: 10.1002/cne.22662; pmid: 21618219
50. E. D. Karagiannis *et al.*, Expansion microscopy of lipid membranes. bioRxiv [Preprint] 829903 (2019); https://doi.org/10.1101/829903.
51. M. R. Tavakoli *et al.*, Light-microscopy-based connectomic reconstruction of mammalian brain tissue. *Nature* **642**, 398–410 (2025). doi: 10.1038/s41586-025-08985-1; pmid: 40335689

ACKNOWLEDGMENTS

We thank J. Zak for discussion on mouse olfactory system and assistance with data analysis; S. Cologna for mouse sample preparation; A. Schantz and D. Keene for human sample collection; J. Sun, W. Cho, and L.-W. Gong for cell cultures; Y. Wang for assistance with the macroscale light-sheet PS experiment; R. Torrez, J. Iwasa, and S. Kwon for assistance with animation and scientific illustration; the Center for Advanced Microscopy & Nikon Imaging Center at Northwestern University and the Chicago Biomedical Consortium (CBC) for access to fluorescence microscopy facilities; the NVIDIA IndeX team for sharing the NVIDIA IndeX software; and J. White for managing the computing cluster at the Advanced Bioimaging Center at UC Berkeley. **Funding:** This work was supported by the National Institutes of Health (grant UG3MH126864 with a subaward to R.G.); the University of Illinois Chicago Startup Fund (R.G.); the Philomathia Foundation (S.U. and E.B.); Lawrence Berkeley National Laboratory's Laboratory Directed Research and Development (LDRD) Program (grants 7647437 and 7721359 to S.U. under US Department of Energy contract no. DE-AC02-05CH11231); the Howard Hughes Medical Institute (E.B.); the Chan Zuckerberg Initiative Imaging Scientist Program (grants 2019-198142 and 2021-244163 to S.U.); Chan Zuckerberg Biohub – San Francisco (investigator gift funding to S.U.); and the Sloan Foundation (grant G-2022-19390 to S.U.). This research used resources of the National Energy Research Scientific Computing Center (NERSC), a US Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory operated under contract no. DE-AC02-05CH11231 using NERSC award DDR-ERCAP0025501. **Author contributions:** Conceptualization: R.G., S.U.; Data curation: X.R., G.L., S.U.; Formal analysis: X.R., S.U., G.L., W.W., R.G.; Funding acquisition: S.U., E.B., R.G.; Investigation: W.W., X.R., G.L.; Methodology: W.W., R.G., X.R., G.L., S.U.; Microscope design, construction, and validation: G.L., D.E.M., S.U., E.B.; Project administration: R.G., S.U.; Resources: W.L., W.W., E.B., S.U., R.G.; Software: X.R., D.E.M., S.U.; Supervision: R.G., S.U.; Validation: W.W., X.R., G.L., S.U., R.G.; Visualization: S.U., R.G., X.R., W.W., G.L.; Writing – original draft: R.G.; Writing – review & editing: W.W., X.R., G.L., D.E.M., W.L., E.B., S.U., R.G. W.W., X.R., and G.L. contributed equally to this work and are listed in reverse alphabetical order. **Competing interests:** R.G. is a coinventor of multiple patents related to expansion microscopy (US Patents 11,873,374 and 12,233,184). The other authors declare no competing interests. Portions of the technology described herein are covered by US Patent 7,894,136 issued to E.B., assigned to Lattice Light of Ashburn, VA, and licensed to Carl Zeiss Microscopy; US Patents 8,711,211 and 9,477,074 issued to E.B., assigned to HHMI, and licensed to Carl Zeiss Microscopy; and US Patent 9,500,846 issued to E.B. Portions of the technology described herein are included in a patent in process (US Patent application no. 63/678,132). **Data and materials availability:** Custom-synthesized materials and reagents are available to researchers within the capacity of the R.G. lab through a simple interinstitutional materials transfer (MTA) agreement (such as a UBMTA). Custom scripts used for petabyte-scale image data processing and analysis ("PetaVIPS") are available on GitHub at https://github.com/abcucberkeley/PetaVIPS. The version associated with this manuscript is available on Zenodo (*36*). The uncompressed raw and processed data for this paper exceed a petabyte in total and thus are not practical to be hosted on currently available public repositories. All data used in this manuscript are available upon request with a feasible data transfer and storage mechanism (such as a Globus endpoint or cloud bucket). All other data are available in the main text or the supplementary materials. **License information:** Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. This article is subject to HHMI's Open Access to Publications policy. HHMI lab heads have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the Author Accepted Manuscript (AAM) of this article can be made freely available under a CC BY 4.0 license immediately upon publication.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adr9109
Materials and Methods; Supplementary Notes; Figs. S1 to S26; Table S1; References (*52*–*63*); Movies S1 and S2; MDAR Reproducibility Checklist

Submitted 30 July 2024; resubmitted 18 July 2025; accepted 17 September 2025
10.1126/science.adr9109

RESEARCH ARTICLE SUMMARY

CELL BIOLOGY

"Kiss-shrink-run" unifies mechanisms for synaptic vesicle exocytosis and hyperfast recycling

Chang-Lu Tao*†, Chong-Li Tian†, Yun-Tao Liu†, Zhen-Hang Lu†, Lei Qi, Xiao-Wei Li, Chao Li, Xuefeng Shen, Min-Ling Gu, Wen-Lan Huang, Shuo Liu, Lei-Qing Yang, Zhenghan Liao, Xiaomin Ma, Jing Wu, Jianyuan Sun, Peiyi Wang, Pak-Ming Lau*, Z. Hong Zhou*, Guo-Qiang Bi*

Full article and list of author affiliations: https://doi.org/10.1126/science.ads7954

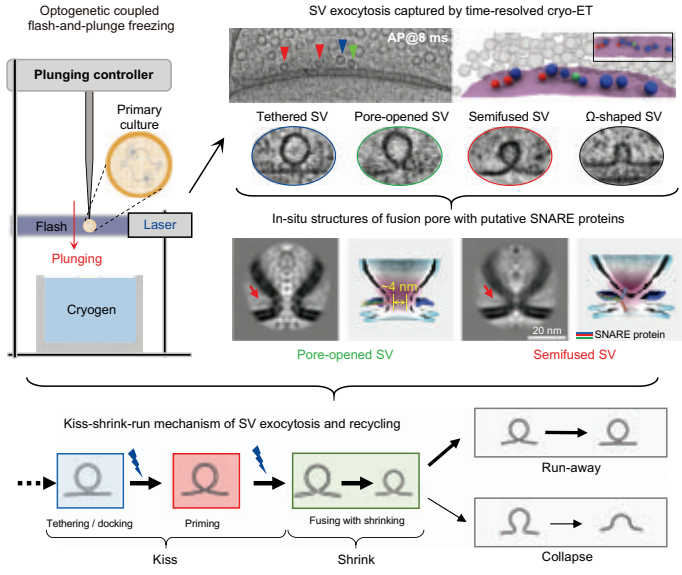
INTRODUCTION: Synaptic vesicle (SV) exocytosis is triggered by an action potential and leads to neurotransmitter release. As such, SV exocytosis is fundamental to neuronal communication. However, the structural and biophysical mechanisms underlying SV exocytosis remain incompletely understood. In particular, questions remain regarding the dynamic interactions between the SV membrane, the presynaptic membrane, and the protein complexes involved. This gap has fueled a long-standing debate over the existence of transient "kiss-and-run" fusion versus irreversible "full-collapse" fusion in central synapses.

RATIONALE: Resolving this debate requires techniques that can achieve nanometer spatial resolution and millisecond temporal resolution. To address this need, we developed a time-resolved, cellular cryo–electron tomography (cryo-ET) method to image intact synapses in cultured rat hippocampal neurons. This approach integrated optogenetic stimulation for synaptic activation and plunge-freezing at millisecond intervals. It also provided high three-dimensional spatial resolution, enabling accurate vesicle size measurements and detailed structural analysis of vesicle–plasma membrane interactions. Using this technique, we acquired more than 1000 tomograms of entire excitatory synapses, frozen at various time points from 0 to 300 ms post–action potential. This large dataset facilitated rigorous statistical analysis of vesicle states and subtomogram averaging to visualize fusion-pore structures.

RESULTS: Near the presynaptic active zones, we observed two distinct SV populations with diameters centered at ~29 and ~41 nm. These SVs were mainly categorized into seven distinct structural states: tethered (large and small), semifused (large and small), pore-opened (large and small), and Ω-shaped (small). Notably, the population of small SVs decreased markedly after neural network inactivity with 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX) and was completely absent when vesicle exocytosis was blocked by tetanus toxin.

Time-resolved cryo-ET revealed a sequence of transitions among these vesicle states. In the resting condition, large SVs were primarily tethered to the plasma membrane (docking). Within 4 ms post–action potential, docked SVs transitioned to large semifused SVs (priming) that transiently "kiss" the plasma membrane. These primed SVs then formed pore-opened SVs with a ~4-nm lipidic fusion pore flanked by putative soluble NSF attachment protein receptor (SNARE) complexes. These fused SVs rapidly shrank to small pore-opened SVs with approximately half the surface area of the original large SVs. Most shrunken SVs subsequently closed their fusion pores and converted into small semifused SVs.

By 70 ms, small semifused SVs detached from the presynaptic membrane ("run-away"), whereas the remaining shrunken, pore-opened SVs fully collapsed into the plasma membrane. After 100 ms, the run-away SVs began to migrate to the periphery of the SV cluster, and the resulting expanded presynaptic membrane began to be retrieved through ultrafast endocytosis.



Kiss-shrink-run mechanism of SV release revealed by time-resolved cryo-ET. A millisecond-precision time-resolved cryo-ET system, involving plunge-freezing coupled with optogenetic stimulation (top left), captured distinct intermediate states of SV exocytosis in intact hippocampal synapses (top right). With population analysis and subtomogram averaging, a kiss-shrink-run sequence was revealed as the prominent SV exocytosis and recycling pathway in hippocampal synapses (bottom). AP, action potential.

CONCLUSION: Our study identifies a kiss-shrink-run pathway as the dominant biophysical mechanism for SV exocytosis and rapid recycling in hippocampal synapses. This kiss-shrink-run mechanism reconciles the kiss-and-run and full-collapse models of neurotransmission and provides a unified explanation for the high efficiency and fidelity of synaptic transmission. Our integrative methodology also establishes a general framework for probing membrane dynamics and molecular interactions in situ with high spatiotemporal precision. □

*Corresponding author. Email: taocl@ustc.edu.cn (C.-L.Ta.); plau@ustc.edu.cn (P.-M.L.); hong.zhou@ucla.edu (Z.H.Z.); gqbi@ustc.edu.cn (G.-Q.B.) †These authors contributed equally to this work. Cite this article as C.-L. Tao *et al.*, *Science* 390, eads7954 (2025). DOI: 10.1126/science.ads7954

CELL BIOLOGY

"Kiss-shrink-run" unifies mechanisms for synaptic vesicle exocytosis and hyperfast recycling

Chang-Lu Tao[1,2,3]*†, Chong-Li Tian[1,2]†, Yun-Tao Liu[1,4,5]†, Zhen-Hang Lu[1,2]†, Lei Qi[1,6], Xiao-Wei Li[1], Chao Li[2], Xuefeng Shen[2], Min-Ling Gu[1,2], Wen-Lan Huang[2], Shuo Liu[2,3], Lei-Qing Yang[2], Zhenghan Liao[2], Xiaomin Ma[7], Jing Wu[7], Jianyuan Sun[2,3], Peiyi Wang[7]‡, Pak-Ming Lau[1,2,6]*, Z. Hong Zhou[4,5]*, Guo-Qiang Bi[1,2,6]*

Synaptic vesicle (SV) exocytosis underpins neuronal communication, yet its nanoscale dynamics remain poorly understood owing to limitations in visualizing rapid events in situ. Here, we used optogenetics-coupled, time-resolved cryo–electron tomography to capture SV exocytosis in rat hippocampal synapses. Within 4 milliseconds of synaptic activation, SVs transiently "kiss" the plasma membrane, forming a ~4-nanometer lipidic fusion pore flanked by putative soluble NSF-attachment protein receptor (SNARE) complexes and then rapidly "shrink" to approximately half of their original surface area. By 70 milliseconds, most shrunken SVs recycle via a "run-away" pathway, whereas others collapse into the presynaptic membrane. Ultrafast endocytosis retrieves the expanded presynaptic membrane after 100 milliseconds. These findings reveal a "kiss-shrink-run" mechanism of SV exocytosis and hyperfast recycling, reconciling conflicting models and elucidating the efficiency and fidelity of synaptic transmission.

The brain's function relies on information transmission between neurons. This task is accomplished by synaptic vesicle (SV) exocytosis, which releases neurotransmitters in response to a presynaptic action potential (AP) (*1*–*5*). Over the past several decades, electrophysiological studies have uncovered a series of rapid events involved in SV exocytosis, including docking, priming, and fusion-pore formation, all of which occur within milliseconds (*5*–*8*). Biochemical and molecular biological analyses have identified key synaptic proteins, particularly the soluble NSF-attachment protein receptor (SNARE) complex, as critical mediators of these biophysical events (*5*, *9*–*12*). However, the fundamental biophysical and structural underpinnings of synaptic transmission—particularly how vesicular and presynaptic membranes and protein complexes interact with one another during this highly dynamic process—have remained elusive (*13*).

[1]Center for Integrative Imaging, Hefei National Research Center for Physical Sciences at the Microscale, MOE Key Laboratory for Membraneless Organelles and Cellular Dynamics, School of Life Sciences, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China. [2]Interdisciplinary Center for Brain Information, the Brain Cognition and Brain Disease Institute, Shenzhen–Hong Kong Institute of Brain Science, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. [3]Faculty of Life Sciences, Shenzhen University of Advanced Technology, Shenzhen, China. [4]California NanoSystems Institute, University of California, Los Angeles, Los Angeles, CA, USA. [5]Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, Los Angeles CA, USA. [6]Anhui Province Key Laboratory of Biomedical Imaging and Intelligent Processing, Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei, China. [7]Cryo-EM Centre, Southern University of Science and Technology, Shenzhen, China. ***Corresponding author. Email: taocl@ustc.edu.cn (C.-L.Ta.); plau@ustc.edu.cn (P.-M.L.); hong.zhou@ucla.edu (Z.H.Z.); gqbi@ustc.edu.cn (G.-Q.B.)** †These authors contributed equally to this work. ‡Present address: Cryo-EM Center, Shandong Agricultural University, Taian, China.

Two main mechanisms of SV exocytosis and recycling have been proposed, the "kiss-and-run" model and the "full-collapse" model. In the kiss-and-run model, the fusion pore opens briefly—allowing neurotransmitter release—then closes, and the vesicle is retrieved for reuse (*14*–*21*). By contrast, the full-collapse fusion model involves the complete merging of the vesicle with the plasma membrane (*3*, *16*). The kiss-and-run model is considered more energy efficient and allows for higher rates of synaptic transmission (*15*, *18*, *22*–*24*).

Classic studies using electron microscopy (EM) have visualized vesicle exocytosis and subsequent endocytosis in the neuromuscular junction at milliseconds and minutes after electrical stimulation, providing direct evidence for the occurrence of full-collapse fusion and clathrin-mediated endocytosis (*3*, *25*). In hippocampal synapses, studies combining optical stimulation with high-pressure freezing EM (flash-and-freeze) revealed the collapse of fused vesicles followed by clathrin-independent endocytosis of large (~80-nm diameter) endocytic vesicles that occur ~100 ms after stimulation (*26*). This process, known as ultrafast endocytosis, has been confirmed through optical imaging and electrophysiological recording experiments (*27*, *28*). Conversely, super-resolution optical imaging has captured the dynamic process of kiss-and-run exocytosis of large dense-core vesicles in chromaffin cells (*29*–*32*). However, the occurrence of kiss-and-run of SVs in central synapses has remained controversial, partly because tools with sufficient temporal and spatial resolution to resolve key features, such as fusion-pore dynamics, are lacking (*13*–*20*, *26*).

In recent years, cryo–electron tomography (cryo-ET) with nanometer spatial resolution has been used to study synaptic ultrastructures, revealing detailed structural features of vesicle interactions with the plasma membrane (*33*–*39*). Particularly, molecular bridges, referred to as "tethers" that connect vesicles with the presynaptic membrane, have been shown to depend on proteins such as SNAP25 or Munc13 (*35*). With high-concentration K^+ spray stimulation followed by plunge-freezing, instances of vesicles directly contacting the plasma membrane have been observed in isolated synaptosomes and interpreted as exocytosis intermediates (*34*). However, owing to limited sample size and possible complications associated with synaptosome preparation, the precise temporal sequence of these intermediates—and whether they represent stages of exocytosis or endocytosis—remains unclear. To directly visualize the dynamic process of SV exocytosis in situ, we developed a time-resolved, cellular cryo-ET method coupled with millisecond-precision optogenetic stimulation.

Distinct subpopulations of SVs near the presynaptic active zone

We imaged intact synapses of rat hippocampal neurons grown on holey carbon-coated gold cryo-EM grids using cellular cryo-ET and obtained their three-dimensional (3D) reconstructions at molecular resolution (Fig. 1, A and B, figs. S1 and S2, and movie S1), following previously established procedures (*38*–*43*). Synapses within the cryo-tomograms were identified based on characteristic features, including closely apposed compartments, one of which was populated by vesicles, and a relatively uniform-width synaptic cleft containing transcleft protein densities (Fig. 1, A to D, figs. S1A and S2A, and movie S1). Synapses were further classified as excitatory or inhibitory based on criteria established in previous correlative imaging studies (*38*, *39*, *41*–*43*). Inhibitory synapses were distinguished by their uniform, thin, sheet-like postsynaptic densities (PSDs) and the presence of γ-aminobutyric acid type A ($GABA_A$) receptor–like densities on postsynaptic membranes (fig. S2A), whereas the remaining synapses were considered excitatory, exhibiting thicker PSDs or occasionally lacking clear PSDs (Fig. 1, A and C). Here, we focused exclusively on excitatory synapses.

Opposite the synaptic cleft and PSD, the presynaptic membrane region known as the active zone (AZ) is where SVs dock and fuse (*44*, *45*). Consistent with the absence of heavy-metal staining in cryo-ET, the dense projection at the AZ that is prominent in conventional

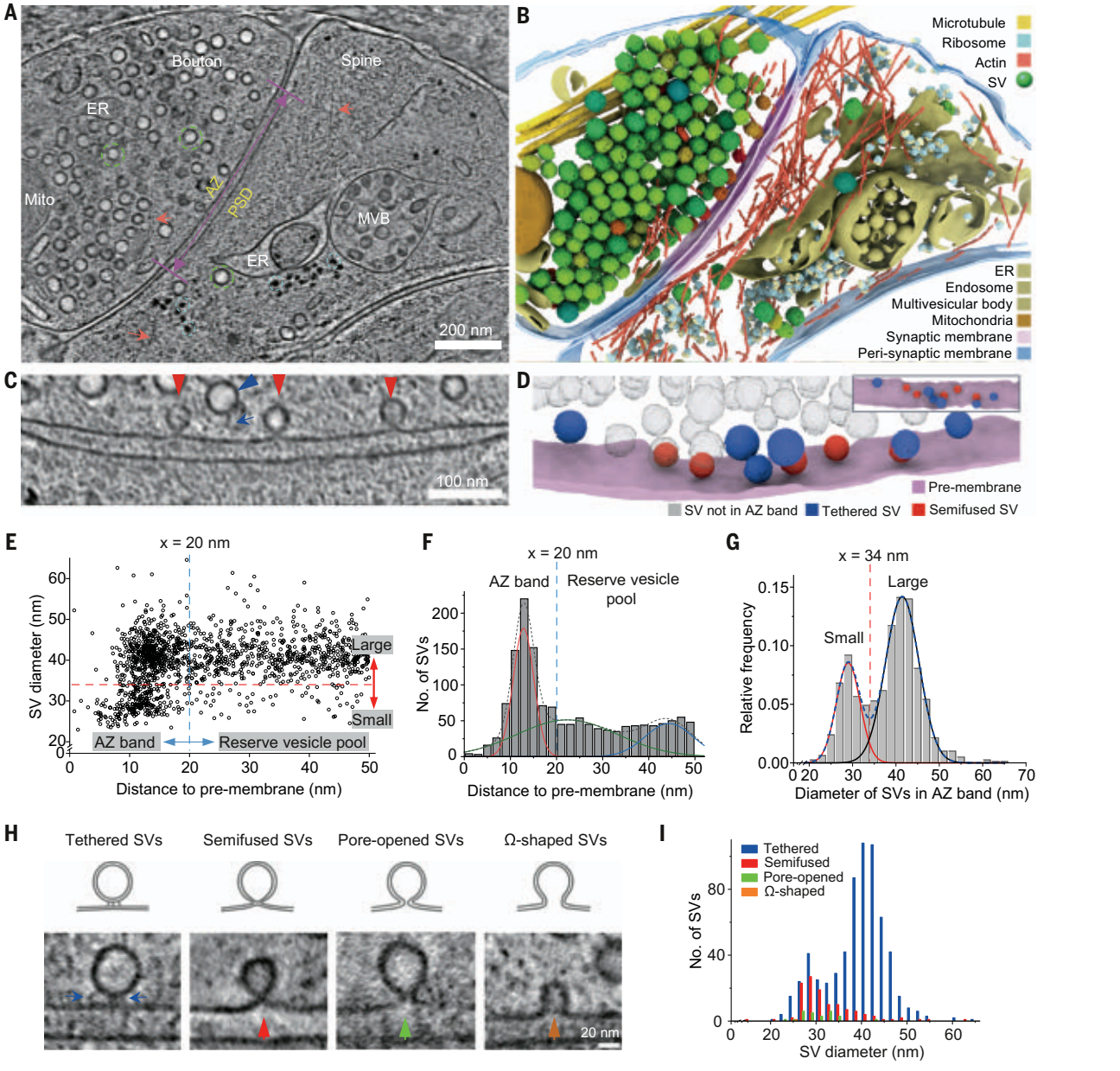


Fig. 1. In situ cryo-ET imaging revealed various types of SVs near the AZs under basal conditions. (**A**) A tomographic slice showing the ultrastructure of an excitatory synapse. Subcellular components—including SVs (green circles), actin filaments (orange arrows), endoplasmic reticulum (ER), mitochondria (Mito), ribosomes (cyan circles), and a multivesicular body (MVB) in the presynaptic bouton and/or postsynaptic spine—are clearly visible. The borders of the presynaptic AZ and PSD are indicated by a magenta double-arrowhead line. (**B**) 3D rendering of the synaptic structures in the tomogram shown in (A). SVs are color-coded by diameter from small (red) to medium (green) to large (blue). (**C**) Zoomed-in view of the AZ from (A) showing tethered vesicles (blue arrows) and semifused vesicles (red arrowheads). (**D**) 3D rendering of the presynaptic profile, including an orthogonal view (inset) of the entire AZ. (**E**) Scatter plot showing vesicle diameter versus distance from the presynaptic membrane for vesicles within 50 nm of the membrane. The red dashed line marks the size threshold distinguishing small from large vesicle clusters, and the blue dashed line denotes the distance threshold separating the AZ band from the reserve vesicle pool. (**F**) Number of SVs as a function of their distances from the presynaptic membrane. The AZ band was defined as the presynaptic region within 20 nm of the AZ (blue dashed line). The red, green, and blue lines represent the three Gaussian components, and the gray dashed line represents their sum. (**G**) Diameter distribution of SVs in the AZ bands, which was fitted with a double-Gaussian function ($\mu1 = 28.99$, $\sigma1 = 2.73$; $\mu2 = 41.46$, $\sigma2 = 3.86$). The intersection point of the two individual Gaussian curves is at 34 nm (red dashed line). The red and black lines indicate the two Gaussian components, and the blue dashed line represents their sum. (**H**) Examples of various types of vesicles in the AZ bands, including tethered, semifused, pore-opened, and Ω-shaped SVs. Colors are the same as in (I). (**I**) Size distributions of different types of SVs in the AZ bands. Statistical analysis combined data from 139 synapses under basal conditions (without CNQX treatment). The number of experiments and data points used for analysis are detailed in table S1.

EM imaging was less apparent in our datasets (*39*, *44*, *46*) (Fig. 1A). We therefore delineated the boundary of the AZ by the region underlying the uniform-width synaptic cleft. We further referred to the ~20-nm cytoplasmic layer immediately beneath the AZ as the submembrane AZ band, which accommodated most membrane-interacting SVs and may correspond to the readily releasable pool (*7*, *23*) (Fig. 1, A to F, and fig. S1A). By quantifying the size of the SVs, we found that many SVs within the AZ bands were substantially smaller than those away from them in the presynaptic boutons (Fig. 1, C to I, and fig. S2). Overall, the size distribution of SVs in the AZ bands exhibited two distinct

Gaussian peaks, one centered at an average diameter of 29 nm and the other at 41 nm; the latter was typical for most of the vesicles located outside the AZ bands (Fig. 1, E and G). We defined 34 nm, the intersection point of the two individual Gaussian curves, as the threshold to categorize SVs into two distinct size groups (Fig. 1G).

Within the AZ band, we identified four major types of SVs: (i) tethered vesicles, linked to the presynaptic membrane via protein filaments; (ii) semifused vesicles, partially fused with the presynaptic membrane; (iii) pore-opened vesicles, which form a narrow fusion pore with the plasma membrane (neck width <22 nm); and (iv) Ω-shaped vesicles, which form a wide fusion pore with plasma membrane (neck width >22 nm) (Fig. 1H and fig. S1D; see methods for details). We also occasionally observed small pits, clathrin-coated vesicles (*47*), and large vesicles resembling those involved in ultrafast endocytosis (*26*) at the presynaptic membrane (fig. S1, F to H). Additionally, a number of tethered SVs were observed at distances up to 50 nm from the presynaptic membrane (fig. S1, I and J). The size of these vesicles strongly correlated with their type: Most tethered SVs had larger diameters, whereas most of the semifused, pore-opened, and Ω-shaped SVs had smaller diameters (Fig. 1I and fig. S1E), indicating that their direct interaction with the plasma membrane may relate to the loss of vesicular membrane.

Action potential–triggered SV exocytosis captured by time-resolved cryo-ET

To investigate whether these different types of SVs in the AZ band reflect different stages of exocytosis, and to determine the chronological sequence of this process, we developed a time-resolved cryo-ET method (Fig. 2 and figs. S3 to S5). In this approach, the action potential (AP) was triggered by optogenetic stimulation, followed by plunge-freezing with millisecond precision (Fig. 2A and fig. S5). We transfected neurons with ChIEF, a variant of channelrhodopsin (*48*), and confirmed that AP firing occurred in these neurons upon light stimulation (Fig. 2, B and C, and fig. S3). To suppress spontaneous recurrent activation of these neurons during sample handling, we applied CNQX (6-cyano-7-nitroquinoxaline-2,3-dione) to block excitatory transmission (*49*).

Optogenetic stimulation was delivered using a fixed-position laser beam intersecting the grid's falling path (Fig. 2A and fig. S5). Using this method, we acquired cryo-tomograms of synapses frozen at approximately 4, 8, 30, and 70 ms after AP firing (post-AP), based on four different fixed positions of the laser beam (Fig. 2, D to I, and fig. S5). The plunge-freezing chamber was set to 34°C so that the ambient temperatures for the 30- and 70-ms positions were 34°C. The ambient temperatures near the cryogen surface were measured at ~24° and ~29°C for the 4- and 8-ms positions, respectively, owing to the proximity of the laser beam (fig. S5, B, C and G; see methods for details). The temporal variability of our time-resolved approach stemmed from three aspects, which were calibrated as follows: (i) The grid's falling speed was consistent, with time variation at any given position of the grid's falling trajectory being less than 0.6 ms, as measured using a fast-speed video recorder (fig. S5, D and E); (ii) the delay of AP induction by optogenetic stimulation was ~4.5 ms and remained temperature independent across 20° to 34°C, calibrated by electrophysiological measurements (fig. S3, D to G); and (iii) synaptic transmission was robustly evoked by optogenetic stimulation across 20° to 34°C, with the kinetics of excitatory postsynaptic current decay slowed at low temperatures (with a variability of 3 ms), calibrated by patch-clamp recordings (fig. S3, H to K). Thus, the minimal temporal variability of our approach ensured confidence in the temporal assignment of the data.

Using this time-resolved cryo-ET approach, we observed substantial increases in the numbers of both the semifused and pore-opened vesicles 4 ms post-AP (Fig. 2, K and L, and fig. S6, A to C), indicating that the generation of these vesicles was triggered by the AP. To investigate whether the formation of semifused and pore-opened SVs involves membrane fusion machinery, we treated the neurons with tetanus neurotoxin (TeNT), which blocks vesicle exocytosis through cleavage of synaptobrevin, a major component of the SNARE complex (*50*) (fig. S4). TeNT completely blocked the formation of semifused and pore-opened vesicles (Fig. 2, I, K, and L) and caused an accumulation of tethered vesicles in both resting and light-stimulated synapses (Fig. 2J). Thus, the AP-triggered formation of semifused and pore-opened vesicles requires functional SNARE complexes.

"Kiss-shrink" process of SV exocytosis

By separating the semifused and pore-opened SVs into large and small subpopulations (Fig. 3A), we observed that the number of large semifused SVs peaked at 4 ms post-AP and then decreased at later time points. By contrast, the numbers of large pore-opened and small pore-opened vesicles peaked at 8 ms, and the number of small semifused SVs continued to increase until 70 ms post-AP (Fig. 3B). These trends indicated a sequence of SV state transition: from large semifused SVs (peak at 4 ms), to large pore-opened SVs (peak at 8 ms), and then quickly to small pore-opened SVs (peak at 8 ms), followed by a slower transition to small semifused SVs (increase until 70 ms) with fusion pore closure. Notably, the large semifused SVs lacked fusion pores and were also observed in resting synapses (Fig. 3, A and B), consistent with the electrophysiologically defined priming state (*6*). In the large semifused SVs, the vesicular membrane made direct contact with a protruded tip of the plasma membrane, forming a metastable intermediate that we refer to as "kissing." After the kissing step, the fusion pore opened, as indicated by an increase in large pore-opened SVs. These vesicles then quickly shrunk into small pore-opened SVs with approximately half of their original surface area, reflected by a diameter reduction from 41 to 29 nm. Thus, we propose a "kiss-shrink" mechanism for vesicle exocytosis: the kissed large semifused vesicles first open the fusion pore and then shrink as neurotransmitter is released (Fig. 3G).

The extensive cryo-ET dataset of SVs interacting with the plasma membrane enabled further analysis of the molecular machinery mediating membrane fusion. We performed subtomogram averaging of both large and small semifused and pore-opened vesicles (Fig. 3, C to F, and fig. S7). This revealed a clear fusion pore, with a defined bilayer wall and a direct tunnel connecting the vesicular lumen to the extracellular space (Fig. 3, D and E). The fusion pore had an approximate diameter of 4 nm, large enough to allow for the release of neurotransmitters, represented here by glutamate, which is less than 1 nm in size. We estimated that neurotransmitter release from these pore-opened vesicles occurs within a time course of ~50 μs, whereas fusion pores likely remain open for about 2 ms (see methods), sufficient for near-complete neurotransmitter release from the vesicle.

In all subtomogram averages of semifused and pore-opened SVs, we identified weak but persistent linear densities flanking the contacting sites (Fig. 3, C to F, and fig. S7). These densities likely correspond to coiled-coil structures of the tripartite SNARE complex, as previously resolved in atomic models, supporting our observation that the formation of semifused and pore-opened SVs was SNARE dependent. Indeed, multiple distinct putative SNARE densities were observed in individual vesicles (fig. S7, A and B). However, reconstructions of these vesicles using different rotational symmetries produced inconsistent results (fig. S7, I and J), suggesting that SNARE protein number and localization may be heterogeneous or that the contact sites are intrinsically asymmetric. Nevertheless, in some subtomogram reconstructions, the structures of synaptotagmin-1 and the SNARE complex could be cofitted (fig. S7J).

Overall, the distinct temporal dynamics of the various SV subpopulations supported a plausible SNARE-dependent kiss-shrink scenario of exocytic events. In the initial kiss step, large semifused vesicles are formed, likely transitioned from tethered vesicles. During the following shrink step, large pore-opened vesicles rapidly transform into small ones. Subsequently, they "pause" at the small semifusion state, as their number continues to increase (Fig. 3G).

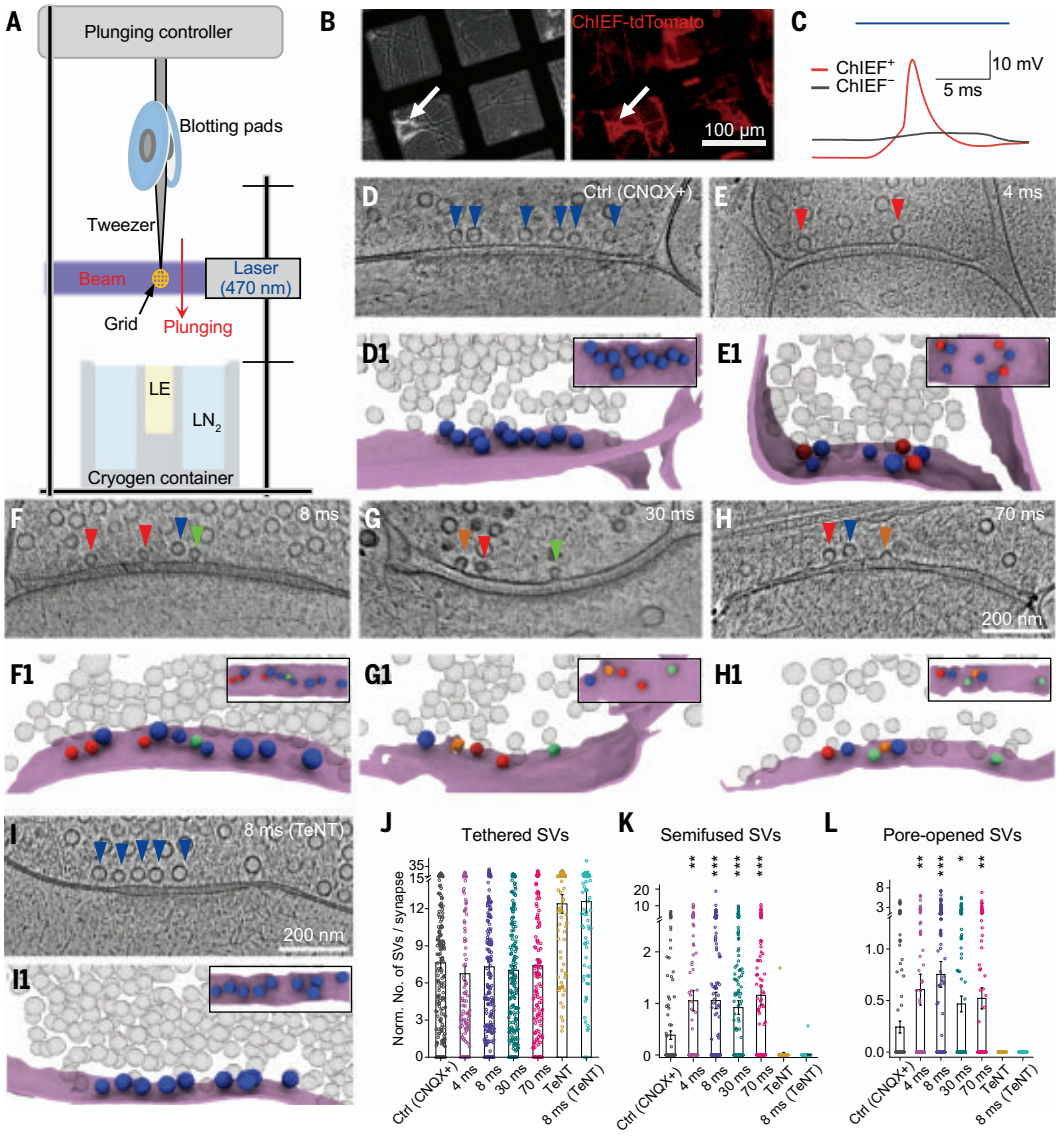


Fig. 2. AP-evoked vesicle exocytosis revealed by millisecond-precision time-resolved cryo-ET. (**A**) Schematic overview of the home-built flash-and-plunge-freezing device. Immediately after blotting with filter papers, the sample grid is plunged into the cryogen [liquid ethane (LE) surrounded by liquid nitrogen (LN_2)]. A 470-nm laser was mounted on the side of the cryo-plunger at a fixed height atop the surface of the liquid ethane to illuminate the dropping grid. (**B**) Bright field (left) and fluorescence (right) microscopy images of hippocampal neurons cultured on an EM grid. The neurons expressed ChIEF-tdTomato proteins (white arrows). (**C**) Patch-clamp recordings showing that APs can be induced in the neuron with ChIEF protein expression ($ChIEF^+$), shown in (B) by the white arrow, using a 470-nm laser stimulation. Under the same light conditions, APs were not evoked in a nearby neuron lacking ChIEF protein expression ($ChIEF^-$). The blue line indicates the duration of light stimulation. (**D** to **I**) Example tomographic slices showing synapses frozen without light stimulation (D); frozen at 4 ms (E), 8 ms (F), 30 ms (G), or 70 ms (H) post-AP; or frozen at 8 ms post-AP under TeNT treatment (I). Arrowheads indicate tethered (blue), semifused (red), pore-opened (green), and Ω-shaped (brown) SVs in the AZ bands. (D1) to (I1) show 3D renderings of the presynaptic profiles in (D) to (I), respectively, including their orthogonal views of the entire AZ (magenta, presynaptic membrane; blue, tethered SVs; red, semifused SVs; green, pore-opened SVs; brown, Ω-shaped SVs; gray, SVs not within the AZ band). (**J** to **L**) Normalized number of tethered (J), semifused (K), and pore-opened (L) SVs in the AZ band of each synapse under different conditions. Data are presented as mean ± SEM. Statistical significance was evaluated using the Mann-Whitney test; $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$. The number of experiments and data points used for analysis are provided in table S1. Specific p values for the significance comparisons are detailed in table S2.

SVs predominantly "run away" after the kiss-shrink step

The accumulation of small semifused SVs after AP stimulation raised the question of their subsequent fate. Notably, in synapses under basal conditions (i.e., without network inactivity by CNQX), the size distribution of SVs that lacked direct contact with the plasma membrane (i.e., tethered and free vesicles) in the AZ band exhibited two distinct peaks at ~29 and ~41 nm, corresponding to small and large SVs, respectively (fig. S6D). By contrast, the small SV peak was substantially reduced after CNQX treatment and completely abolished by TeNT treatment, suggesting that the formation of small tethered and small free SVs depends on synaptic activity and is likely a result of kiss-shrink exocytosis (fig. S6, E, G, and H).

More directly, the number of small tethered SVs in the AZ band tightly correlated with the time elapsed after the AP (Fig. 4, A to C, and fig. S6F). The number of small tethered vesicles began to increase ~30 ms post-AP and continued to rise beyond 70 ms, indicating that these vesicles appear after the small semifused SVs (Fig. 4C). These observations suggested a "run-away" pathway after

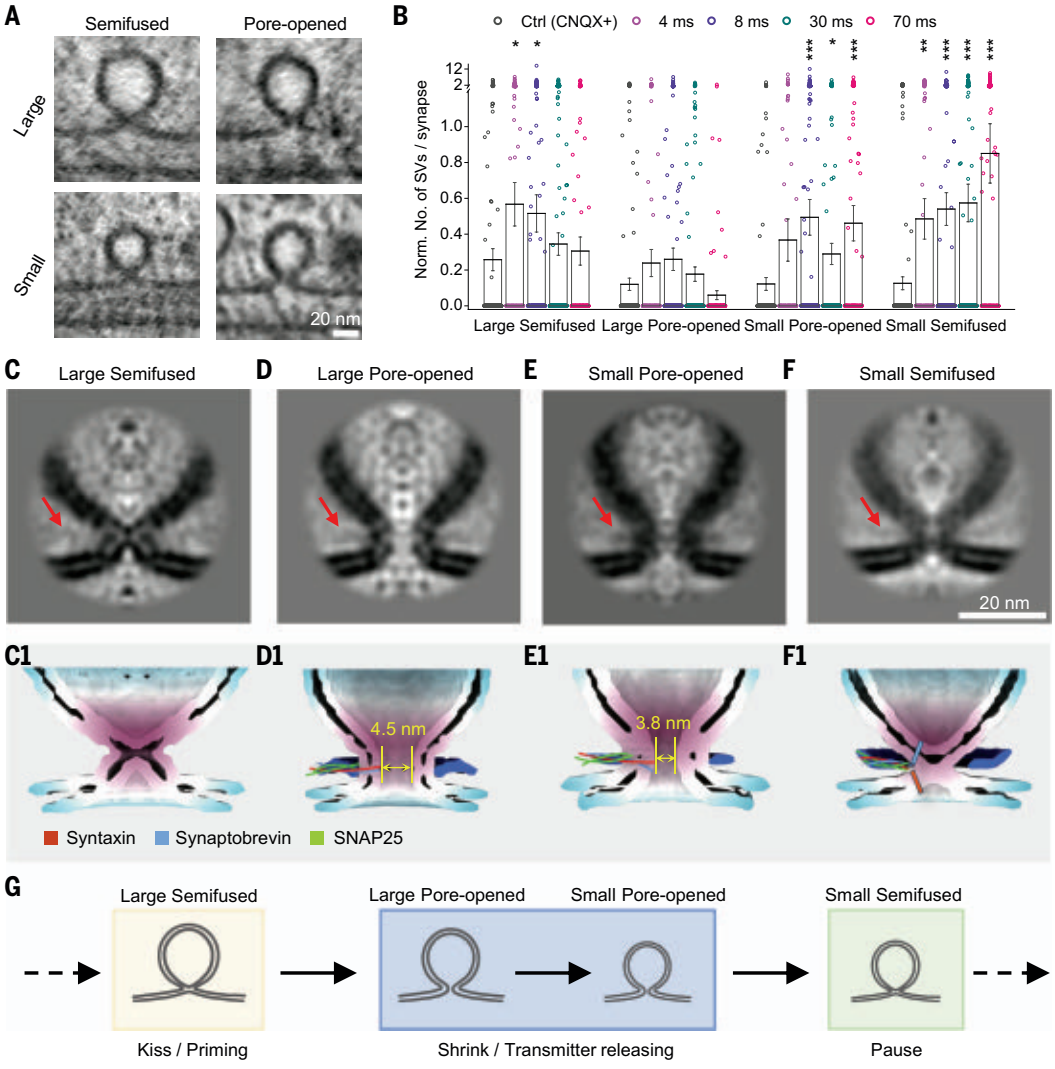


Fig. 3. Kiss-shrink-pause pathway of SV exocytosis. (**A**) Example tomographic slices showing semifused (left) and pore-opened (right) SVs of large (top row) and small (bottom row) sizes. (**B**) Normalized number of the four subtypes of SVs per synapse at different time points post-AP. (**C** to **F**) Central slices of the subtomogram averages for the four subtypes of SVs. Red arrows indicate the putative SNARE complex. (C1) to (F1) present 3D renderings of the subtomogram averages shown in (C) to (F), respectively. Atomic models of the SNARE protein coiled coils [Protein Data Bank (PDB) ID 3HD7 in (D1) and (E1); PDB ID 1SFC in (F1)] are presented as cylinder diagrams, fitted into the cryo-ET density. (**G**) Schematic diagram illustrating the kiss-shrink model of vesicle exocytosis. Data are presented as mean ± SEM. Statistical significance was evaluated using the Mann-Whitney test; $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$. The number of synapses, vesicles, and specific p values for these comparisons are provided in table S3.

kiss-shrink fusion, thus completing a "kiss-shrink-run" route of exocytosis and endocytosis.

In addition, we observed vesicles collapsing into the presynaptic membrane, exhibiting Ω-shaped or pitlike profiles, during later stages of AP-triggered exocytosis (Fig. 4, D and E). Most of these collapsing SVs were small, comparable to the small semifused or pore-opened SVs in size (Fig. 4F). Chronologically, most of the collapsing SVs appeared around 70 ms post-AP, after the increase in small pore-opened vesicles, suggesting a "kiss-shrink-collapse" sequence. At all time points after the AP, the number of collapsing SVs remained a small fraction of the total vesicles compared with those following the run-away path (Fig. 4G). Thus, the kiss-shrink-run pathway appears to be the dominant mechanism of SV exocytosis and recycling in hippocampal synapses.

Capturing the entire process of kiss-shrink-run and kiss-shrink-collapse allowed us to correlate morphologically defined large and small tethered, semifused, and pore-opened vesicles with the previously defined docking and priming vesicle states (*10*, *35*, *51*) (Fig. 4H). A subpopulation of large tethered vesicles with very short tethers may correspond to the docking state. Following this, the distinct large semifused vesicles can be considered as being in a priming state, ready to open fusion pores. Notably, the small vesicles (including small pore-opened, small semifused, small tethered, and small free SVs), which accounted for about ~20% of the SVs in the AZ band and up to 70% of the SVs directly contacting the plasma membrane, are in the postrelease state and no longer in the docking or priming states.

Run-away vesicle migration and ultrafast endocytosis after 100 ms post-AP

To further investigate the fate of run-away small vesicles in the AZ band (i.e., small tethered and small free SVs), we performed a time-resolved cryo-ET study on synapses frozen at >100 ms post-AP (fig. S8, A and B). At this time point, we observed small vesicles located outside of the AZ band (Fig. 5, A and B). We suspected that these vesicles

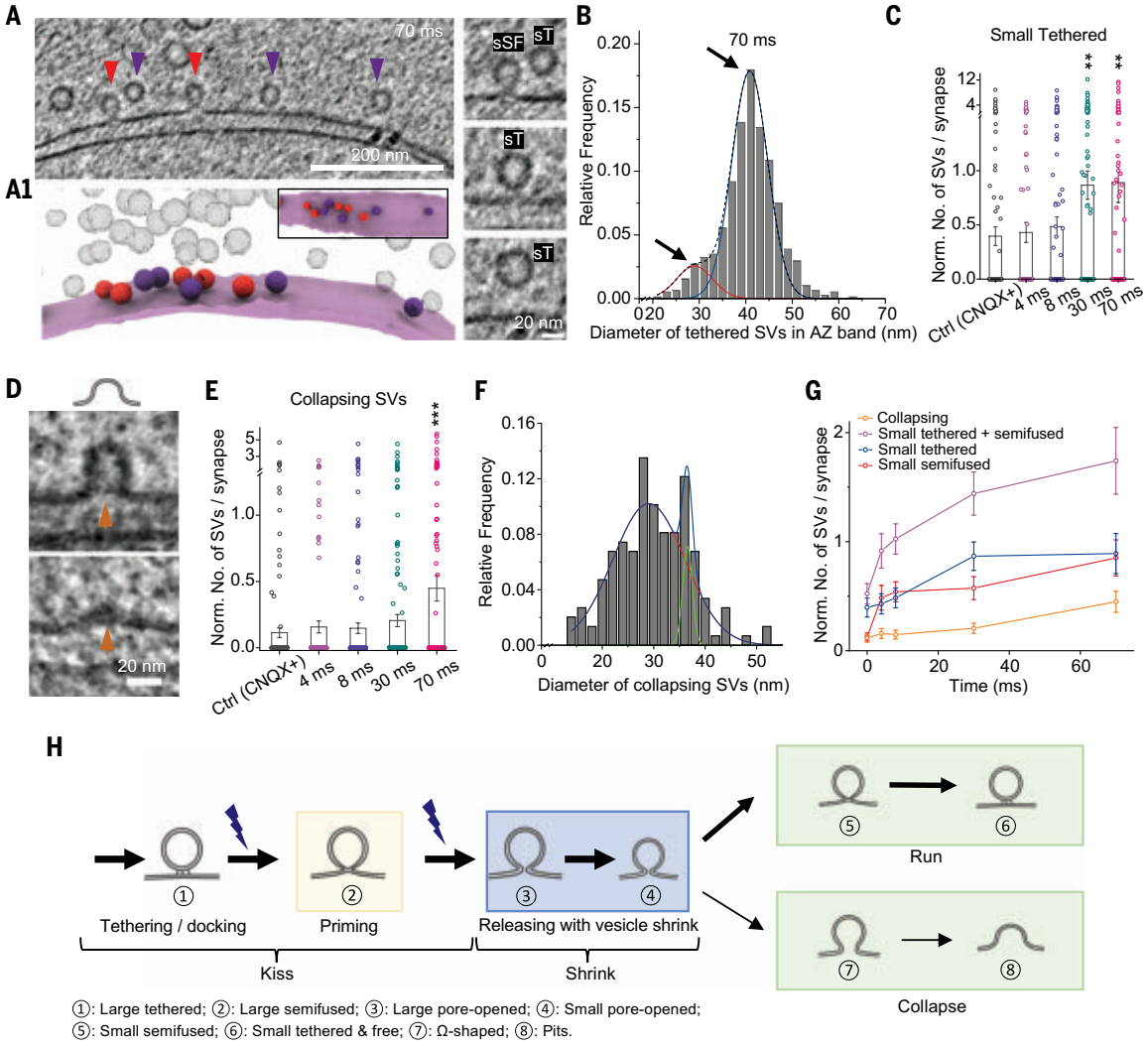
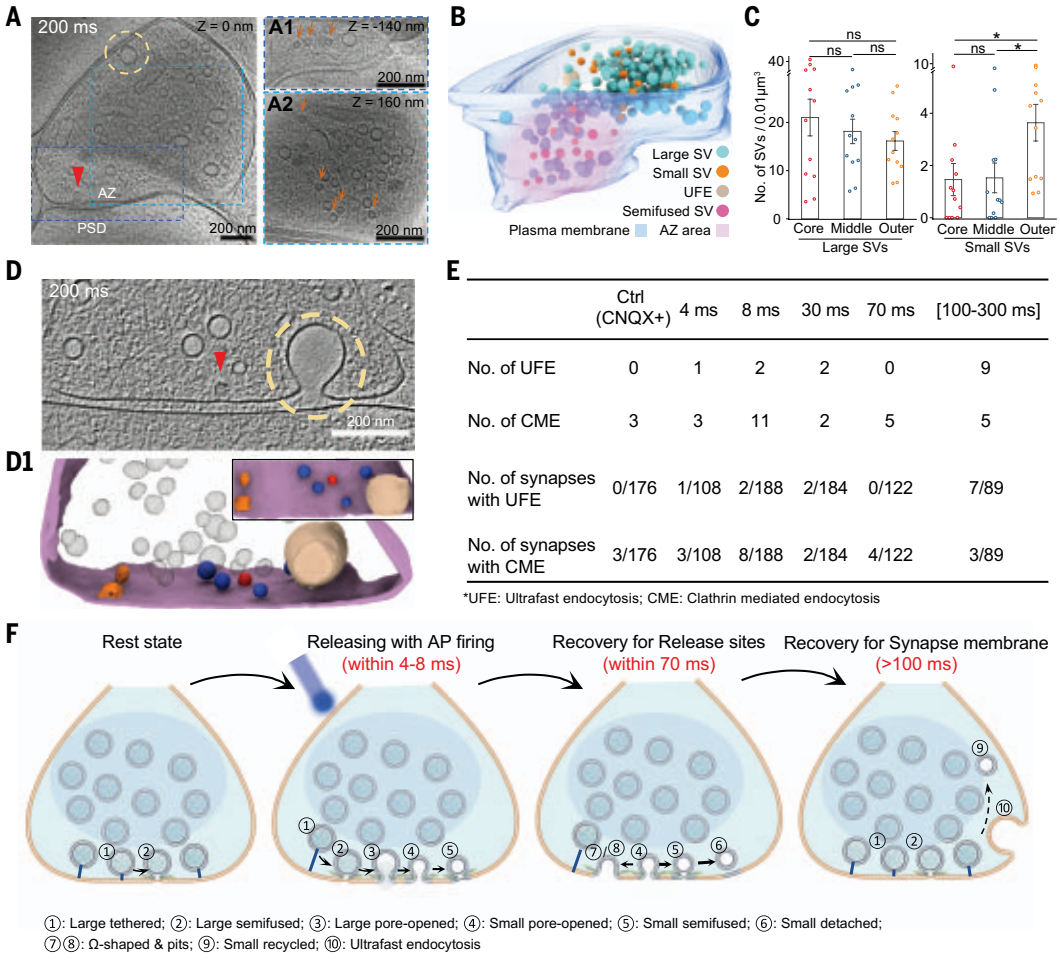


Fig. 4. Vesicle recycling predominantly occurs through the run-away pathway after kiss-shrink exocytosis. (**A**) Tomographic slice (left) of a synapse frozen at 70 ms post-AP, showing small tethered SVs (purple arrowheads) and small semifused SVs (red arrowheads) within the AZ band. Shown on the right are zoomed-in views of individual SVs in the AZ band, categorized as small semifused (sSF) and small tethered (sT) SVs. (A1) presents a 3D rendering of the presynaptic profile, including the orthogonal view of the entire AZ band (magenta, presynaptic membrane; red, semifused SVs; purple, small tethered SVs; gray, SVs not in the AZ band). (**B**) Size distribution of tethered SVs in the AZ bands of synapses frozen at 70 ms post-AP, revealing two peaks (black arrows). The diameter distribution was fitted with a double-Gaussian function. The red and blue lines represent two Gaussian components, and the black dashed line represents their sum. (**C**) Normalized number of small tethered SVs in the AZ bands of synapses frozen at different time points post-AP. (**D**) Example tomographic slices of collapsing SVs, including Ω-shaped SVs (pore neck width >22 nm; top) and pits (bottom). (**E**) Normalized number of collapsing SVs per synapse at different time points post-AP. (**F**) Diameter distribution of collapsing SVs, fitted with a double-Gaussian function ($\mu1 = 29.06$, $\sigma1 = 5.959$; $\mu2 = 36.57$, $\sigma2 = 0.79$). The red and green lines represent two Gaussian components, and the blue line represents their sum. (**G**) Normalized number of different types of SVs in the AZ bands of synapses frozen at different time points post-AP. (**H**) Schematic diagram depicting the two concurrent routes of vesicle exocytosis in hippocampal synapses: the kiss-shrink-run and the kiss-shrink-collapse. Arrows indicate transitions between vesicle states, with the arrow thickness roughly corresponding to the relative rate of transition. The blue lightning bolts indicate synaptic activations triggered with optogenetic stimulation. Data are presented as mean ± SEM. Statistical significance was determined using the Mann-Whitney test; $^{**}p < 0.01$, and $^{***}p < 0.001$. The number of synapses and vesicles, along with p values for significance comparisons, are provided in table S4.

migrated from the AZ band after the kiss-shrink-run pathway. These recycled SVs were primarily localized at the periphery of the vesicle cluster within the presynaptic bouton (Fig. 5C and fig. S8C), suggesting that run-away vesicles do not integrate directly into the reserve pool of regular-sized vesicles.

An alternative explanation for the appearance of small vesicles outside the AZ band is that they may result from peri-AZ release. To evaluate this possibility, we analyzed synapses frozen at 0 to 300 ms post-AP and found that exocytic events outside the AZ were fewer than 1 of 10 of those occurring within the AZ (fig. S8, D, E, and G). By contrast, the number of small free vesicles outside the AZ band was found to be five times higher (fig. S8F), implying that most of these vesicles had migrated from the AZ band. Although we cannot entirely exclude minor peri-AZ release, its contribution appears minimal.

We also observed an increase in large endocytic structures (~80 nm in diameter) in the peri-AZ region at 100 to 300 ms post-AP (Fig. 5, A, D, and E), likely corresponding to previously reported ultrafast endocytosis (*26*). By contrast, clathrin-mediated endocytosis events remained unchanged across the 0- to 300-ms time span (Fig. 5E), likely because clathrin-mediated endocytosis occurs over a much longer timescale (seconds to minutes). These observations suggested that ultrafast endocytosis could compensate for membrane loss resulting



	Ctrl (CNQX+)	4 ms	8 ms	30 ms	70 ms	[100-300 ms]
No. of UFE	0	1	2	2	0	9
No. of CME	3	3	11	2	5	5
No. of synapses with UFE	0/176	1/108	2/188	2/184	0/122	7/89
No. of synapses with CME	3/176	3/108	8/188	2/184	4/122	3/89

*UFE: Ultrafast endocytosis; CME: Clathrin mediated endocytosis

Fig. 5. Recycling of shrunken vesicles and presynaptic membrane after 100 ms post-AP. (**A**) Tomographic slice of a synapse frozen at 200 ms post-AP, showing one semifused SV at the AZ (red arrowhead) and one large vesicle undergoing ultrafast endocytosis at a peri-AZ site (yellow dashed circle). (A1) and (A2) show zoomed-in views of the boxed areas in (A) taken at different Z heights through the top (A1) and bottom (A2) of the presynaptic bouton. Small vesicles (brown arrows) were enriched at the top and bottom of the vesicle pool. (**B**) 3D rendering of all vesicles in the entire presynaptic bouton shown in (A). (**C**) The density of large (left) and small (right) vesicles across three layers of the entire vesicle pool, excluding vesicles in the AZ band. (**D**) Tomographic slice of a synapse frozen at 200 ms post-AP, showing a large vesicle under ultrafast endocytosis at the edge of the AZ (yellow dashed circle). (D1) presents a 3D rendering of the presynaptic profile, including the orthogonal view of the entire AZ band (magenta, presynaptic membrane; red, semifused SVs; blue, tethered SVs; yellow, collapsing SVs; gray, SVs not in the AZ band). (**E**) Number of vesicles undergoing ultrafast endocytosis (UFE) and clathrin-mediated endocytosis (CME) in synapses under different conditions. (**F**) Schematic diagram depicting the organization and dynamic of SVs in the presynaptic bouton during AP-induced synaptic release. In the resting condition, SVs are either tethered to the presynaptic membrane or form large semifused vesicles, ready for release in response to AP firing. Within ~4 ms post-AP, approximately one vesicle per synapse rapidly releases its transmitter through the kiss-shrink pathway. By 70 ms post-AP, most shrunken SVs detach from the presynaptic membrane (run-away), whereas others collapse into the presynaptic membrane. After 100 ms post-AP, the run-away shrunken vesicles further migrate to the periphery of the vesicle pool, and the expanded presynaptic membrane begins recycling through ultrafast endocytosis. Data are shown as mean ± SEM. Statistical significance was determined using the Mann-Whitney test; *$p < 0.05$, and ns indicates $p > 0.05$. The number of vesicles, along with p values for significance comparisons, are provided in table S5.

from both the kiss-shrink-run and kiss-shrink-collapse modes of vesicle exocytosis (Fig. 5F).

Discussion

By capturing transient intermediate states of SV exocytosis with millisecond precision, we have demonstrated that SVs undergo a rapid kiss-shrink-run process upon AP firing (Figs. 4H and 5F and movie S2). In the sequence, a SV first makes direct contact with the presynaptic membrane (kiss), then rapidly shrinks to approximately one-third of their original volume after forming a narrow fusion pore, releasing neurotransmitters through both diffusion and pressure-driven expulsion (shrink). The SV then typically detaches (run) within tens of milliseconds, enabling hyperfast recycling (Fig. 4H). This kiss-shrink-run pathway accounts for more than 80% of releasing events, whereas a smaller fraction of vesicles undergoes a kiss-shrink-collapse type of exocytosis. Conceivably, the lost vesicular membrane during both kiss-shrink-run and kiss-shrink-collapse events must be retrieved into the presynaptic bouton, likely through subsequent endocytosis over different timescales, including the well-documented ultrafast endocytosis within a few hundred milliseconds post-AP and perhaps slower clathrin-mediated endocytosis (*19*, *25*–*28*, *52*, *53*). Thus, our findings reconcile the historically debated kiss-and-run and full-collapse models and provide a unified mechanistic framework for the process of vesicle exocytosis and recycling (Fig. 5F and movie S2).

At the beginning of the kiss-shrink-run process, AP stimulation apparently facilitates the transition of a large tethered vesicle into a

metastable, large semifused state, which may correspond to the previously described priming (*5*, *7*, *10*). This semifused vesicle is then further triggered by AP-induced Ca^{2+} influx, leading to the opening of the fusion pore that bridges the vesicular and plasma membrane and facilitates two concurrent processes. First, neurotransmitters within the vesicle's lumen diffuse out through the fusion pore. Second, the vesicle undergoes rapid shrinkage to roughly one-third of its original volume, which further accelerates the diffusion of its contents. Overall, the kiss-shrink process facilitates neurotransmitter release from SVs.

Mechanistically, the driving force for vesicle shrinkage may arise from membrane tension imbalance after fusion-pore formation (*29*, *54*–*59*), which typically promotes pore dilation and presumably causes the collapse of the vesicle into the plasma membrane. However, our data showed that the fusion pore can be rather stably maintained. This suggested the existence of a constricting barrier surrounding the fusion pore, likely formed by SNARE complexes and/or their regulatory proteins, which mediates the interactions between vesicular and plasma membranes (*9*, *36*, *60*, *61*). Moreover, the size of a shrunken vesicle also stabilized at about half of its original membrane area, probably reflecting the balance of factors such as membrane curvature, molecular composition, protein interaction, and restricted diffusion.

After the kiss-shrink phase, the run process retrieves the shrunken vesicle with its vesicular proteins within tens of milliseconds. This not only provides a hyperfast route for membrane recycling but also leaves a "clean" site for a new vesicle to dock and fuse. Thus, the kiss-shrink-run mechanism enables the hyperfast turnover of vesicles at the active zone, supporting high rates of synaptic transmission (*24*). The shrunken SV likely contains a more concentrated set of proteins and possesses a lipid composition that favors higher membrane curvature (*55*, *62*). Consequently, upon eventual detachment from the presynaptic membrane, the shrunken vesicle may be unable to dock and fuse again. Indeed, these recycled vesicles typically did not integrate into the reserve SV pool in the presynaptic bouton. In this way, the kiss-shrink-run process may act as a natural quality control mechanism, preventing neurotransmitter-free vesicles from occupying limited number of docking sites and responding to APs (Fig. 5F). Thus, in addition to improving the efficiency of vesicle recycling and recovery of release sites, the kiss-shrink-run route ensures the fidelity of synaptic transmission.

The uncovering of the kiss-shrink-run pathway relied on our new strategy—time-resolved cryo-ET of intact hippocampal synapses—which offered three key advantages: (i) sufficient temporal resolution of optogenetic stimulation and plunge-freezing, enabling synchronization with synaptic activation; (ii) high spatial resolution in 3D, allowing precise size measurements of individual vesicles and detailed analysis of structural features of their interaction with the plasma membrane, for example, the narrow fusion pores; and (iii) a sufficiently large sample size of more than 1000 intact synapses, ensuring statistically rigorous analyses. By combining these advantages, we were able to quantitatively analyze a series of exocytic events and the hyperfast recycling of small vesicles. More generally, this integrative strategy allows in situ investigation of cellular membrane dynamics and molecular interactions with high spatiotemporal precision.

Is kiss-shrink-run a general mechanism? In theory, it is not impossible that vesicle exocytosis and recycling in various types of synapses may occur along different biophysical or cellular pathways. Thus, whereas cultures of dissociated hippocampal neurons provide valuable assays for studying fundamental biophysical mechanisms such as SV dynamics, our findings should be interpreted with caution. Notably, datasets from previous studies using high-pressure freezing, freeze substitution, and room-temperature ET of acute brain slices or slice culture (*10*, *63*, *64*), as well as studies using cryo-ET imaging of isolated synaptosomes undergoing high-concentration K^+ spray-mixing plunge-freezing (*34*), suggest that docked or primed vesicles may exist in different sizes. It would be informative to quantitatively analyze whether they also belong to distinct vesicle states. Future studies with more physiological preparations, for example, by combing flash-freezing of brain slices with cryo-ET, will provide insight into the broader physiological relevance of the kiss-shrink-run mechanism and its role in synaptic communication across various neuronal types.

Materials and methods are available in the supplementary materials.

REFERENCES AND NOTES

1. J. del Castillo, B. Katz, Quantal components of the end-plate potential. *J. Physiol.* **124**, 560–573 (1954). doi: 10.1113/jphysiol.1954.sp005129; pmid: 13175199
2. J. del Castillo, B. Katz, Local activity at a depolarized nerve-muscle junction. *J. Physiol.* **128**, 396–411 (1955). doi: 10.1113/jphysiol.1955.sp005315; pmid: 14392617
3. J. E. Heuser *et al.*, Synaptic vesicle exocytosis captured by quick freezing and correlated with quantal transmitter release. *J. Cell Biol.* **81**, 275–300 (1979). doi: 10.1083/jcb.81.2.275; pmid: 38256
4. T. C. Südhof, J. Rizo, Synaptic vesicle exocytosis. *Cold Spring Harb. Perspect. Biol.* **3**, a005637 (2011). doi: 10.1101/cshperspect.a005637; pmid: 22026965
5. T. C. Südhof, Neurotransmitter release: The last millisecond in the life of a synaptic vesicle. *Neuron* **80**, 675–690 (2013). doi: 10.1016/j.neuron.2013.10.022; pmid: 24183019
6. U. Becherer, J. Rettig, Vesicle pools, docking, priming, and release. *Cell Tissue Res.* **326**, 393–407 (2006). doi: 10.1007/s00441-006-0243-z; pmid: 16819626
7. E. Neher, N. Brose, Dynamically primed synaptic vesicle states: Key to understand synaptic short-term plasticity. *Neuron* **100**, 1283–1291 (2018). doi: 10.1016/j.neuron.2018.11.024; pmid: 30571941
8. H. Taschenberger, A. Woehler, E. Neher, Superpriming of synaptic vesicles as a common basis for intersynapse variability and modulation of synaptic strength. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E4548–E4557 (2016). doi: 10.1073/pnas.1606383113; pmid: 27432975
9. M. Bera *et al.*, Synaptophysin chaperones the assembly of 12 SNAREpins under each ready-release vesicle. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2311484120 (2023). doi: 10.1073/pnas.2311484120; pmid: 37903271
10. C. Imig *et al.*, The morphological and molecular nature of synaptic vesicle priming at presynaptic active zones. *Neuron* **84**, 416–431 (2014). doi: 10.1016/j.neuron.2014.10.009; pmid: 25374362
11. A. T. Brunger, U. B. Choi, Y. Lai, J. Leitz, Q. Zhou, Molecular mechanisms of fast neurotransmitter release. *Annu. Rev. Biophys.* **47**, 469–497 (2018). doi: 10.1146/annurev-biophys-070816-034117; pmid: 29792815
12. A. T. Brunger *et al.*, The pre-synaptic fusion machinery. *Curr. Opin. Struct. Biol.* **54**, 179–188 (2019). doi: 10.1016/j.sbi.2019.03.007; pmid: 30986753
13. N. Brose *et al.*, Synaptic vesicle fusion: Today and beyond. *Nat. Struct. Mol. Biol.* **26**, 663–668 (2019). doi: 10.1038/s41594-019-0277-z; pmid: 31384060
14. N. L. Chanaday, M. A. Cousin, I. Milosevic, S. Watanabe, J. R. Morgan, The synaptic vesicle cycle revisited: New insights into the modes and mechanisms. *J. Neurosci.* **39**, 8209–8216 (2019). doi: 10.1523/JNEUROSCI.1158-19.2019; pmid: 31619489
15. A. A. Alabi, R. W. Tsien, Perspectives on kiss-and-run: Role in exocytosis, endocytosis, and neurotransmission. *Annu. Rev. Physiol.* **75**, 393–422 (2013). doi: 10.1146/annurev-physiol-020911-153305; pmid: 23245563
16. N. C. Harata, A. M. Aravanis, R. W. Tsien, Kiss-and-run and full-collapse fusion as modes of exo-endocytosis in neurosecretion. *J. Neurochem.* **97**, 1546–1570 (2006). doi: 10.1111/j.1471-4159.2006.03987.x; pmid: 16805768
17. L. G. Wu, E. Hamid, W. Shin, H. C. Chiang, Exocytosis and endocytosis: Modes, functions, and coupling mechanisms. *Annu. Rev. Physiol.* **76**, 301–331 (2014). doi: 10.1146/annurev-physiol-021113-170305; pmid: 24274740
18. L. He, L. G. Wu, The debate on the kiss-and-run fusion at synapses. *Trends Neurosci.* **30**, 447–455 (2007). doi: 10.1016/j.tins.2007.06.012; pmid: 17765328
19. S. P. Gandhi, C. F. Stevens, Three modes of synaptic vesicular recycling revealed by single-vesicle imaging. *Nature* **423**, 607–613 (2003). doi: 10.1038/nature01677; pmid: 12789331
20. C. F. Stevens, J. H. Williams, "Kiss and run" exocytosis at hippocampal synapses. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 12828–12833 (2000). doi: 10.1073/pnas.230438697; pmid: 11050187
21. B. Ceccarelli, W. P. Hurlbut, A. Mauro, Turnover of transmitter and synaptic vesicles at the frog neuromuscular junction. *J. Cell Biol.* **57**, 499–524 (1973). doi: 10.1083/jcb.57.2.499; pmid: 4348791
22. S. O. Rizzoli, R. Jahn, Kiss-and-run, collapse and 'readily retrievable' vesicles. *Traffic* **8**, 1137–1144 (2007). doi: 10.1111/j.1600-0854.2007.00614.x; pmid: 17645434
23. A. A. Alabi, R. W. Tsien, Synaptic vesicle pools and dynamics. *Cold Spring Harb. Perspect. Biol.* **4**, a013680 (2012). doi: 10.1101/cshperspect.a013680; pmid: 22745285
24. E. Neher, What is rate-limiting during sustained synaptic activity: Vesicle supply or the availability of release sites. *Front. Synaptic Neurosci.* **2**, 144 (2010). doi: 10.3389/fnsyn.2010.00144; pmid: 21423530
25. J. E. Heuser, T. S. Reese, Evidence for recycling of synaptic vesicle membrane during transmitter release at the frog neuromuscular junction. *J. Cell Biol.* **57**, 315–344 (1973). doi: 10.1083/jcb.57.2.315; pmid: 4348786

26. S. Watanabe *et al.*, Ultrafast endocytosis at mouse hippocampal synapses. *Nature* **504**, 242–247 (2013). doi: 10.1038/nature12809; pmid: 24305055
27. N. L. Chanaday, E. T. Kavalali, Optical detection of three modes of endocytosis at hippocampal synapses. *eLife* **7**, e36097 (2018). doi: 10.7554/eLife.36097; pmid: 29683423
28. I. Delvendahl, N. P. Vyleta, H. von Gersdorff, S. Hallermann, Fast, temperature-sensitive and clathrin-independent endocytosis at central synapses. *Neuron* **90**, 492–498 (2016). doi: 10.1016/j.neuron.2016.03.013; pmid: 27146271
29. P. J. Wen *et al.*, Actin dynamics provides membrane tension to merge fusing vesicles into the plasma membrane. *Nat. Commun.* **7**, 12604 (2016). doi: 10.1038/ncomms12604; pmid: 27576662
30. W. D. Zhao *et al.*, Hemi-fused structure mediates and controls fusion and fission in live cells. *Nature* **534**, 548–552 (2016). doi: 10.1038/nature18598; pmid: 27309816
31. L. Ge *et al.*, Sequential compound fusion and kiss-and-run mediate exo- and endocytosis in excitable cells. *Sci. Adv.* **8**, eabm6049 (2022). doi: 10.1126/sciadv.abm6049; pmid: 35714180
32. W. Shin *et al.*, Visualization of membrane pore in live cells reveals a dynamic-pore theory governing fusion and endocytosis. *Cell* **173**, 934–945.e12 (2018). doi: 10.1016/j.cell.2018.02.062; pmid: 29606354
33. R. G. Held, J. Liang, A. T. Brunger, Nanoscale architecture of synaptic vesicles and scaffolding complexes revealed by cryo-electron tomography. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2403136121 (2024). doi: 10.1073/pnas.2403136121; pmid: 38923992
34. J. Radecke *et al.*, Morphofunctional changes at the active zone during synaptic vesicle exocytosis. *EMBO Rep.* **24**, e55719 (2023). doi: 10.15252/embr.202255719; pmid: 36876590
35. C. Papantoniou *et al.*, Munc13- and SNAP25-dependent molecular bridges play a key role in synaptic vesicle priming. *Sci. Adv.* **9**, eadf6222 (2023). doi: 10.1126/sciadv.adf6222; pmid: 37343100
36. A. Radhakrishnan *et al.*, Symmetrical arrangement of proteins under release-ready vesicles in presynaptic terminals. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2024029118 (2021). doi: 10.1073/pnas.2024029118; pmid: 33468631
37. R. Fernández-Busnadiego *et al.*, Cryo-electron tomography reveals a critical role of RIM1α in synaptic vesicle tethering. *J. Cell Biol.* **201**, 725–740 (2013). doi: 10.1083/jcb.201206063; pmid: 23712261
38. Y. T. Liu *et al.*, Mesophasic organization of $GABA_A$ receptors in hippocampal inhibitory synapses. *Nat. Neurosci.* **23**, 1589–1596 (2020). doi: 10.1038/s41593-020-00729-w; pmid: 33139942
39. C. L. Tao *et al.*, Differentiation and characterization of excitatory and inhibitory synapses by cryo-electron tomography and correlative microscopy. *J. Neurosci.* **38**, 1493–1510 (2018). doi: 10.1523/JNEUROSCI.1548-17.2017; pmid: 29311144
40. Y. T. Liu, C. L. Tao, Digitalizing neuronal synapses with cryo-electron tomography and correlative microscopy. *Curr. Opin. Neurobiol.* **76**, 102595 (2022). doi: 10.1016/j.conb.2022.102595; pmid: 35797809
41. R. Sun *et al.*, An efficient protocol of cryo-correlative light and electron microscopy for the study of neuronal synapses. *Biophys. Rep.* **5**, 111–122 (2019). doi: 10.1007/s41048-019-0092-4
42. Y. T. Liu, C. L. Tao, P. M. Lau, Z. H. Zhou, G. Q. Bi, Postsynaptic protein organization revealed by electron microscopy. *Curr. Opin. Struct. Biol.* **54**, 152–160 (2019). doi: 10.1016/j.sbi.2019.02.012; pmid: 30904821
43. C. L. Tao, Y. T. Liu, Z. H. Zhou, P. M. Lau, G. Q. Bi, Accumulation of dense core vesicles in hippocampal synapses following chronic inactivity. *Front. Neuroanat.* **12**, 48 (2018). doi: 10.3389/fnana.2018.00048; pmid: 29942253
44. T. C. Südhof, The presynaptic active zone. *Neuron* **75**, 11–25 (2012). doi: 10.1016/j.neuron.2012.06.012; pmid: 22794257
45. R. Couteaux, M. Pécot-Dechavassine, [Synaptic vesicles and pouches at the level of "active zones" of the neuromuscular junction]. *C. R. Acad. Sci. Hebd. Seances Acad. Sci. D* **271**, 2346–2349 (1970). pmid: 4995202
46. R. Fernández-Busnadiego *et al.*, Quantitative analysis of the native presynaptic cytomatrix by cryoelectron tomography. *J. Cell Biol.* **188**, 145–156 (2010). doi: 10.1083/jcb.200908082; pmid: 20065095
47. H. T. McMahon, E. Boucrot, Molecular mechanism and physiological functions of clathrin-mediated endocytosis. *Nat. Rev. Mol. Cell Biol.* **12**, 517–533 (2011). doi: 10.1038/nrm3151; pmid: 21779028
48. J. Y. Lin, M. Z. Lin, P. Steinbach, R. Y. Tsien, Characterization of engineered channelrhodopsin variants with improved properties and kinetics. *Biophys. J.* **96**, 1803–1814 (2009). doi: 10.1016/j.bpj.2008.11.034; pmid: 19254539
49. P. M. Lau, G. Q. Bi, Synaptic mechanisms of persistent reverberatory activity in neuronal networks. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10333–10338 (2005). doi: 10.1073/pnas.0500717102; pmid: 16006530
50. G. Schiavo *et al.*, Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin. *Nature* **359**, 832–835 (1992). doi: 10.1038/359832a0; pmid: 1331807
51. G. F. Kusick *et al.*, Synaptic vesicles transiently dock to refill release sites. *Nat. Neurosci.* **23**, 1329–1338 (2020). doi: 10.1038/s41593-020-00716-1; pmid: 32989294
52. S. Watanabe, Slow or fast? A tale of synaptic vesicle recycling. *Science* **350**, 46–47 (2015). doi: 10.1126/science.aad2996; pmid: 26430111
53. J. Balaji, T. A. Ryan, Single-vesicle imaging reveals that synaptic vesicle exocytosis and endocytosis are coupled by a single stochastic mode. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 20576–20581 (2007). doi: 10.1073/pnas.0707574105; pmid: 18077369
54. J. Preobraschenski *et al.*, Synaptophysin accelerates synaptic vesicle fusion by expanding the membrane upon neurotransmitter loading. *Sci. Adv.* **11**, eads4661 (2025). doi: 10.1126/sciadv.ads4661; pmid: 40267188
55. K. L. Budzinski *et al.*, Large structural change in isolated synaptic vesicles upon loading with neurotransmitter. *Biophys. J.* **97**, 2577–2584 (2009). doi: 10.1016/j.bpj.2009.08.032; pmid: 19883601
56. M. M. Kozlov, L. V. Chernomordik, Membrane tension and membrane fusion. *Curr. Opin. Struct. Biol.* **33**, 61–67 (2015). doi: 10.1016/j.sbi.2015.07.010; pmid: 26282924
57. J. C. Shillcock, R. Lipowsky, Tension-induced fusion of bilayer membranes and vesicles. *Nat. Mater.* **4**, 225–228 (2005). doi: 10.1038/nmat1333; pmid: 15711550
58. J. H. Kim *et al.*, Mechanical tension drives cell membrane fusion. *Dev. Cell* **32**, 561–573 (2015). doi: 10.1016/j.devcel.2015.01.005; pmid: 25684354
59. T. T. Kliesch *et al.*, Membrane tension increases fusion efficiency of model membranes in the presence of SNAREs. *Sci. Rep.* **7**, 12070 (2017). doi: 10.1038/s41598-017-12348-w; pmid: 28935937
60. J. Zhu *et al.*, Synaptotagmin rings as high-sensitivity regulators of synaptic vesicle docking and fusion. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2208337119 (2022). doi: 10.1073/pnas.2208337119; pmid: 36103579
61. J. E. Rothman, S. S. Krishnakumar, K. Grushin, F. Pincet, Hypothesis - buttressed rings assemble, clamp, and release SNAREpins for synaptic transmission. *FEBS Lett.* **591**, 3459–3480 (2017). doi: 10.1002/1873-3468.12874; pmid: 28983915
62. K. Komorowski *et al.*, Neurotransmitter uptake of synaptic vesicles studied by x-ray diffraction. *Eur. Biophys. J.* **51**, 465–482 (2022). doi: 10.1007/s00249-022-01609-w; pmid: 35904588
63. C. Imig *et al.*, Ultrastructural imaging of activity-dependent synaptic membrane-trafficking events in cultured brain slices. *Neuron* **108**, 843–860.e8 (2020). doi: 10.1016/j.neuron.2020.09.004; pmid: 32991831
64. K. Lichter *et al.*, Ultrastructural analysis of wild-type and RIM1α knockout active zones in a large cortical synapse. *Cell Rep.* **40**, 111382 (2022). doi: 10.1016/j.celrep.2022.111382; pmid: 36130490
65. C.-L. Tao *et al.*, "Kiss-shrink-run" unifies mechanisms for synaptic vesicle exocytosis and hyperfast recycling. Zenodo (2025); https://doi.org/10.5281/zenodo.16846019.

ACKNOWLEDGMENTS

We thank P. Tang for technical advice and support on cryo-EM imaging; H. Wang, F. Xu, and C. Wang for help with the optogenetic setup; D. Shi and A. Chai for help with electrophysiology experiments; and T. Nguyen and D. Bi for editing the manuscript. We acknowledge the use of instruments at the Center for Integrative Imaging of Hefei National Research Center for Physical Sciences at the Microscale of USTC, the Cryo-EM Centre of Southern University of Science and Technology, and the Electron Imaging Center for Nanomachines of the University of California, Los Angeles (UCLA). We also acknowledge support from Shenzhen Brain Science Infrastructure. **Funding:** This work was funded by STI2030-Major Projects grant 2022ZD0205900 (C.-L.Ta.), National Natural Science Foundation of China (NSFC) grant 32494760 (G.-Q.B.), NSFC grant 32461160291 (G.-Q.B.), Chinese Academy of Sciences grant XDB32030212 (G.-Q.B.), NSFC grant 31630030 (G.-Q.B.), NSFC grant 32200784 (C.-L.Ta.),NSFC grant 32521003 (C.-L.Ta.), NSFC grant 31621002 (G.-Q.B.), Guangdong Zhujiang Talent Program – Young Top Talents project 2021QN02Y964 (C.-L.Ta.), Shenzhen Fundamental Research Program grant JCYJ20210324141013033 (C.-L.Ta.), STI2030-Major Projects grant 2022ZD0211900 (J.W.), Key-Area Research and Development Program of Guangdong Province grant 2023B0303010001 (G.-Q.B.), Guangdong Basic and Applied Basic Research Foundation grant 2021A1515110281 (L.-Q.Y.), and National Institutes of Health grant R01GM071940 (Z.H.Z.). **Author contributions:** Conceptualization: C.-L.Ta., Y.-T.L., G.-Q.B.; Methodology: C.-L.Ta., C.-L.Ti., Y.-T.L., Z.-H.L., L.Q., X.S., P.W., J.S., Z.H.Z., G.-Q.B.; Investigation: C.-L.Ta., C.-L.Ti., Y.-T.L., Z.-H.L., L.Q., X.-W.L., C.L., M.-L.G., W.-L.H., S.L., L.-Q.Y., Z.L., X.M., J.W.; Visualization: C.-L.Ta., C.-L.Ti., Y.-T.L., Z.-H.L., S.L., G.-Q.B.; Funding acquisition: C.-L.Ta., L.-Q.Y., J.W., Z.H.Z., G.-Q.B.; Project administration: C.-L.Ta., G.-Q.B.; Supervision: C.-L.Ta., P.-M.L., Z.H.Z., G.-Q.B.; Writing – original draft: C.-L.Ta., C.-L.Ti., Y.-T.L., Z.-H.L., X.-W.L., C.L., W.-L.H., P.-M.L., Z.H.Z., G.-Q.B.; Writing – review & editing: C.-L.Ta., C.-L.Ti., Y.-T.L., Z.-H.L., L.Q., P.-M.L., Z.H.Z., G.-Q.B. **Competing interests:** The University of Science and Technology of China has filed a patent application related to the design of Flash-and-Plunge Freezing device, on which C.-L.Ta., C.-L.Ti., Y.-T.L., Z.-H.L., and G.-Q.B. are named inventors. The remaining authors declare no competing interests. **Data and materials availability:** Represented subtomogram average maps of the semifused and pore-opened vesicles have been deposited in the Electron Microscopy Data Bank (EMDB) under accession nos. EMD-65182, EMD-65183, EMD-65184, and EMD-65186. Custom-written codes can be found at Zenodo (*65*). **License information:** Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

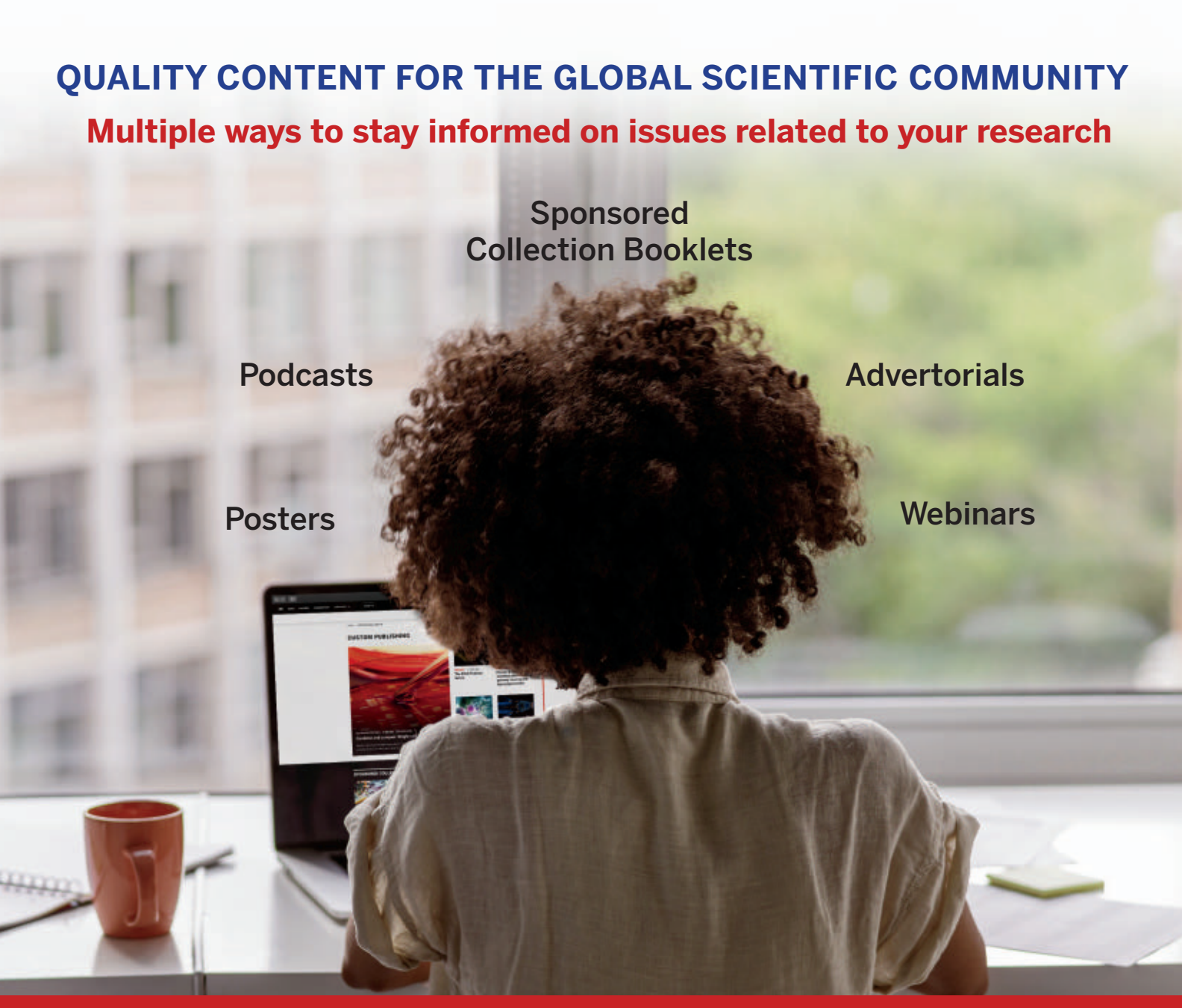
SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ads7954
Materials and Methods; Figs. S1 to S8; Tables S1 to S8; References (*66*–*82*); MDAR Reproducibility Checklist; Movies S1 and S2

Submitted 7 September 2024; resubmitted 9 July 2025; accepted 29 August 2025
10.1126/science.ads7954

QUALITY CONTENT FOR THE GLOBAL SCIENTIFIC COMMUNITY

Multiple ways to stay informed on issues related to your research



Science
AAAS

Scan the code and start exploring the latest advances in science and technology innovation!

Science.org/custom-publishing

Brought to you by the *Science*/AAAS Custom Publishing Office.

Posters **Podcasts** **Sponsored Collection Booklets** **Advertorials** **Webinars**

RESEARCH ARTICLE SUMMARY

NEUROSCIENCE

Head-direction cells as a neural compass in bats navigating outdoors on a remote oceanic island

Shaked Palgi†, Saikat Ray†, Shir R. Maimon†, Yuval Waserman, Liron Ben-Ari, Tamir Eliav, Avishag Tuval, Chen Cohen, Julius D. Keyyu, Abdalla I. Ali, Henrik Mouritsen, Liora Las*, Nachum Ulanovsky*

Full article and list of author affiliations: https://doi.org/10.1126/science.adw6202

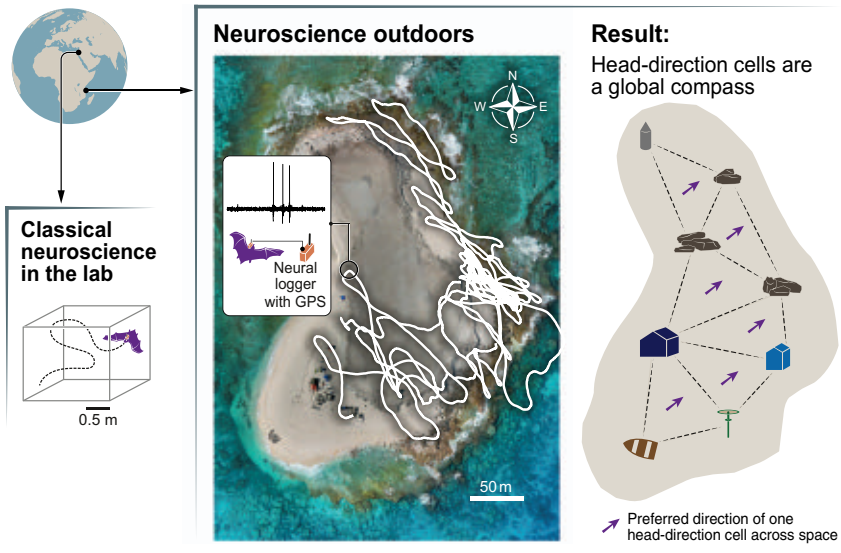
INTRODUCTION: Animals and humans live in large and rich environments and need to navigate to find their way. Navigation behavior outdoors has been studied extensively by ecologists and ethologists. By contrast, the neural mechanisms of navigation have almost exclusively been studied indoors, in small laboratory settings, which are much smaller and contain poorer sensory information than real-world outdoor environments. This leaves a major gap in our understanding of the neural basis of navigation: How does the brain's "navigation circuit" operate in real-world conditions? In this study, we focused on head-direction cells—neurons that change their activity depending on an animal's orientation and are often called "neural compasses." These neurons are well characterized in small environments but had not previously been studied in natural environments outdoors.

RATIONALE: We set out to study head-direction cells in the wild by releasing Egyptian fruit bats to fly freely and navigate unconstrained on a small island near Zanzibar in East Africa. We developed a miniaturized wireless recording device—a "neural logger"—that allowed tracking of the bats' position and direction using a precise GPS and simultaneously recording the activity of multiple individual neurons. We conducted neural recordings in brain regions that are well known to contain head-direction cells. We sought to differentiate between two competing hypotheses: the "mosaic hypothesis," which predicts that each neuron will show different directional tuning in different areas of the island, depending on local sensory cues, and the "global compass hypothesis," which predicts that each neuron will exhibit the same directional tuning across the entire geographical extent of the island. Additionally, manipulating distant visual cues in laboratory experiments causes head-direction tuning to rotate, leading to a potential conundrum: Do head-direction cells rotate their tuning outdoors upon movement of the Moon and stars, which are the most prominent distant visual cues outdoors? If so, then these neurons would form a highly unreliable neural compass for real-world navigation.

RESULTS: We discovered head-direction cells in bats navigating on the island, which had similar functional properties to classical indoor head-direction cells. These neurons maintained their directional tuning over a large geographical scale, consistent with the global compass hypothesis. In contrast, there was no compelling evidence for the mosaic hypothesis. We further showed that head-direction cells remained stable both when the Moon was below the horizon and when the Moon and stars were occluded by clouds, which implies that the geographically stable tuning did not depend on the availability of these distant visual cues. We also ruled out a magnetic field–based origin of the neural compass. Finally, we found that the head-direction code stabilized slowly over several nights, suggesting a process during which bats gradually learned the layout of their environment, including the landmarks and geometry of the island.

CONCLUSION: Our results suggest that head-direction cells represent direction stably across geographical scales and irrespective of celestial dynamics and can therefore serve as the brain's neural compass. This study emphasizes the need to investigate the neural mechanisms of navigation in the wild. More generally, we call for broadening the scope of brain research toward neuroscience in the real world. □

*Corresponding author. Email: nachum.ulanovsky@weizmann.ac.il (N.U.); liora.las@weizmann.ac.il (L.L.) †These authors contributed equally to this work. **Cite this article as** S. Palgi *et al.*, *Science* **390**, eadw6202 (2025). DOI: 10.1126/science.adw6202



Head-direction cells recorded outdoors form a reliable neural compass for real-world navigation. We performed single-unit neural recordings in animals navigating unconstrained in the wild—in bats flying outdoors on a remote oceanic island—and discovered that head-direction cells stably encode the animals' orientation across the entire island, irrespective of the dynamics of the Moon and stars. [Illustration: Genia Brodsky; Photograph: Shaked Palgi]

NEUROSCIENCE

Head-direction cells as a neural compass in bats navigating outdoors on a remote oceanic island

Shaked Palgi[1]†, Saikat Ray[1]†, Shir R. Maimon[1]†, Yuval Waserman[1], Liron Ben-Ari[1], Tamir Eliav[1]‡, Avishag Tuval[1], Chen Cohen[1], Julius D. Keyyu[2], Abdalla I. Ali[3], Henrik Mouritsen[4], Liora Las[1]*, Nachum Ulanovsky[1]*

Animals and humans rely on their navigation skills to survive. However, spatial neurons in the brain's "navigation circuit" had not previously been studied under real-world conditions. We conducted an electrophysiological study of spatial neurons in the wild: We recorded head-direction cells from the presubiculum of bats flying unconstrained and navigating outdoors on a remote oceanic island. These neurons represented the bats' orientation stably across the island's entire geographical scale and irrespective of the dynamics of the Moon and the Milky Way. The directional tuning stabilized over several nights from the first exploration of the island. These results imply that head-direction cells can serve as a learned, reliable neural compass for real-world navigation—highlighting the power of taking neuroscience out into the wild.

Navigation is crucial for animals and humans. Historically, studies of navigation followed two very different approaches: Ethologists and ecologists tracked animals in the wild, focusing on navigational strategies and sensory mechanisms in real-world environments (*1*–*6*). Sensory mechanisms were also investigated behaviorally in laboratory settings (*7*–*9*). In parallel, neuroscientists recorded neural activity from animals navigating indoors in small laboratory enclosures and discovered place cells (*10*, *11*), grid cells (*12*, *13*), and head-direction cells (*14*–*22*)—neurons that are considered to be important for navigation. However, the assumption that the brain's navigation circuit behaves the same in the lab and in the wild is challenged by an increasing body of literature demonstrating that the use of naturalistic stimuli often yields unexpected findings about brain function (*23*). In nature, animals navigate in large open environments with multiple sets of cues, and the landmarks visible to them change as they traverse through space (*24*), which is very different from the typical laboratory setups in which navigation-related neurons are usually recorded (*25*, *26*). These neurons had never been recorded during real-world navigation outdoors, leaving a major gap in our current understanding of the neural mechanisms of navigation: Are these spatial neurons relevant to the real world? We thus studied neurons in the brain's navigation circuit during outdoor navigation—in bats flying freely on an island—focusing on the head-direction circuit.

Head-direction cells are neurons that are tuned to the animal's orientation. In laboratory experiments, these cells were shown in some cases to rotate their tuning after manipulation of the surrounding environment (*16*, *17*, *27*–*30*) and in other cases to remain stable despite such manipulation (*30*–*33*)—depending on the details of the experiment. It thus remains unknown how head-direction cells function in natural environments: Do they encode direction locally, by having a mosaic of several representations anchored to different landmarks at different locations [fig. S1, left; see also (*34*)]? Or do they remain stable and encode direction over regional or global scales, maintaining the same preferred direction across large geographical areas (fig. S1, right)?

Outdoors, some of the most prominent distal visual cues are celestial cues, such as the Sun and the Moon, which are used for navigation by many animals (*35*–*41*). However, these are not stable directional cues, because they appear, disappear, and move in the sky. How do head-direction cells deal with this instability? In laboratory experiments, head-direction tuning rotates when distal cues are rotated (*17*, *27*). On the basis of these studies, we may expect that head-direction tuning outdoors shifts with the daily movement of the celestial cues. This presents a conundrum: Can head-direction cells serve as a reliable neural compass for outdoor navigation? If not, how do animals maintain a stable representation of direction during real-world navigation?

To tackle this challenge, we developed a new recording system that combines wireless neural logging and GPS tracking and conducted single-cell neural recordings in bats flying outdoors. We asked three questions: (i) Do head-direction cells exist over geographical scales outdoors? (ii) Are they local or global compasses over such spatial scales? (iii) Are these neural compasses stable and robust to the movement of celestial cues? We recorded head-direction cells from the dorsal presubiculum of Egyptian fruit bats (*Rousettus aegyptiacus*) as the bats navigated outdoors on Latham Island, a small island near East Africa. We focused on the dorsal presubiculum because head-direction cells in this region have been studied extensively in rodents and bats (*15*, *16*, *22*, *25*). Experiments on an island allowed us to overcome two major obstacles for unconstrained neural recordings outdoors: how to recapture the bats, and how to obtain good behavioral coverage to allow reliable estimation of neuronal tuning.

Conducting single-unit neural recordings in the presubiculum of bats flying outdoors

First, we developed a miniaturized neural logger, which includes a high-precision GPS module together with additional sensors (Fig. 1A). This allowed us to combine precise localization outdoors (Fig. 1, B and C; positional localization accuracy, SD = 71 cm; directional accuracy, SD = 3.3°) with single-cell neural recordings in the bat presubiculum (Fig. 1, D and E, and fig. S2A).

We captured wild bats (n = 6) in mainland Tanzania (Fig. 1F) (*42*) and implanted them with a 16-tetrode microdrive. We then sailed with the implanted animals to Latham Island (Fig. 1, F to H; figs. S3, A and B, and S4, A to C; and Movie 1)—a barren and uninhabited remote island, ~350 m by 250 m in size—where we stayed for 3 to 4 weeks. On the island, we first accustomed the bats to the environment for a few days (inside a cage and a flight tent) and then started nightly recordings during which we released the bats one by one to fly outdoors (the bats were recaptured at the end of each session). The bats were released primarily from two platforms, where food was available (Fig. 1I and green dots in Fig. 1G) (*42*).

During the sessions, which typically lasted ~30 to 50 min (Fig. 1J, top), the bats flew fast (Fig. 1K), at an altitude of several meters (Fig. 1L), and exhibited complex trajectories (Fig. 1G and fig. S2B), allowing them to cover a large area each night (Fig. 1M and fig. S3C; median area covered per session = 5751 m^2; see fig. S5 for additional behavioral parameters). Because of occlusions from obstacles such as cliffs, boulders, and tents, the bats could see only a fraction of the island at any given moment (Fig. 1N and fig. S6). The bats performed many turns (Fig. 1O), resulting in full coverage of the azimuthal (horizontal) axis of heading direction (Fig. 1, P and Q, and fig. S2C). Head direction and heading (flight) direction were highly correlated (Fig. 1R,

[1]Department of Brain Sciences, Weizmann Institute of Science, Rehovot, Israel. [2]Tanzania Wildlife Research Institute, Arusha, Tanzania. [3]Department of Natural Sciences, The State University of Zanzibar, Zanzibar, Tanzania [4]Research Centre for Neurosensory Sciences, and Institut für Biologie und Umweltwissenschaften, Carl-von-Ossietzky Universität Oldenburg, Oldenburg, Germany. ***Corresponding author. Email: nachum.ulanovsky@weizmann.ac.il (N.U.); liora.las@weizmann.ac.il (L.L.)** †These authors contributed equally to this work. ‡Present address: Max Planck Institute for Biological Intelligence, Seewiesen, Germany.

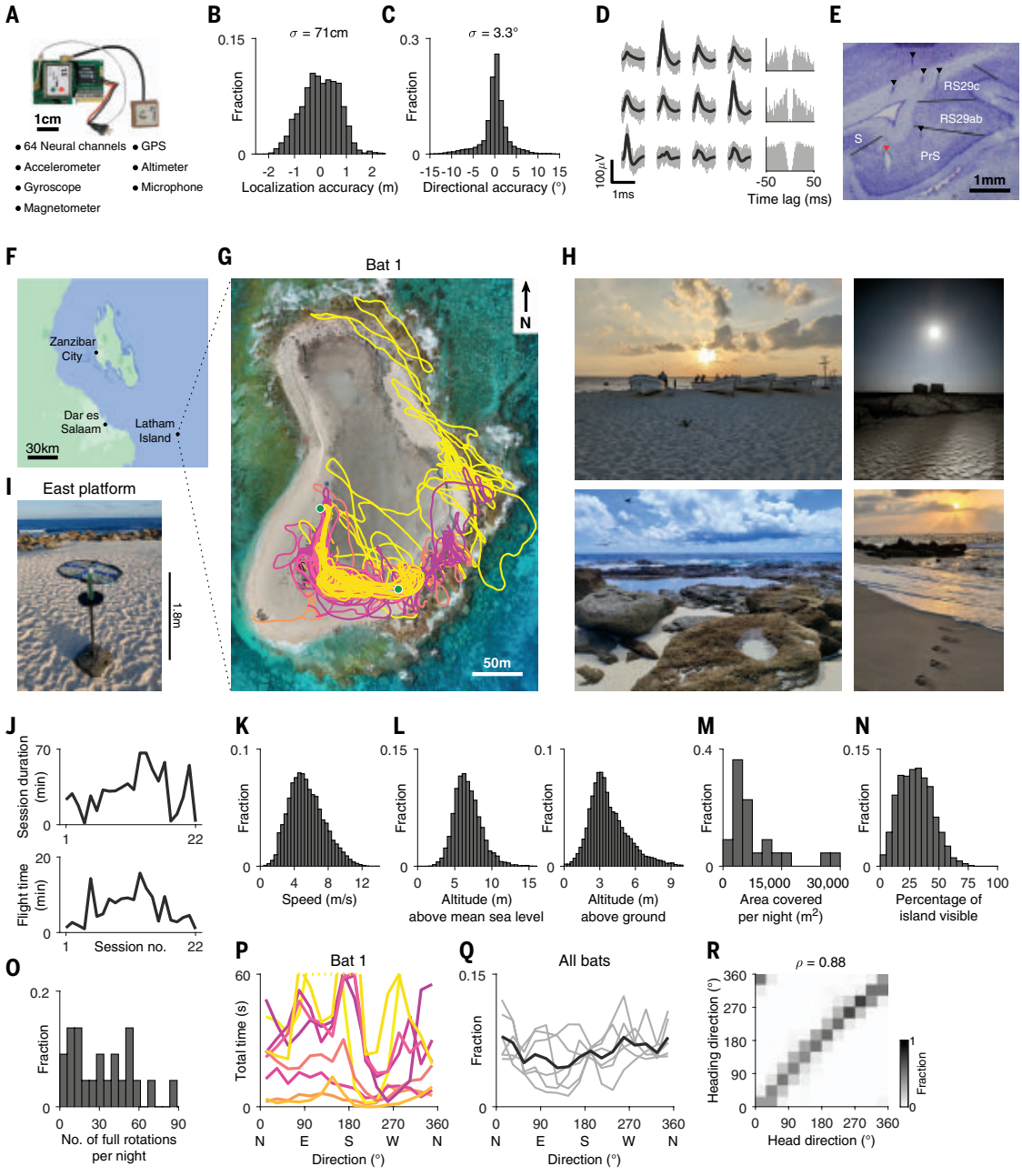


Fig. 1. Conducting single-unit neural recordings in the presubiculum of bats flying outdoors. (**A**) Photograph of the neural logger with GPS, and a list of the onboard sensors. (**B**) GPS localization accuracy, measured in a separate calibration procedure [$n = 26{,}416$ samples (*42*)]; σ, standard deviation = 71 cm. (**C**) GPS directional accuracy [same as (B)]; $\sigma = 3.3°$. (**D**) Three well-isolated neurons (rows) recorded outdoors on one tetrode. (Left) The four channels of the tetrode (columns), showing mean waveforms (black) and 100 random waveforms per neuron (gray). (Right) Spike time autocorrelations, demonstrating clear refractory periods. (**E**) Nissl-stained coronal section of bat 1, showing tetrode tracks toward the dorsal presubiculum (black arrowheads) and an electrolytic lesion in the presubiculum (red arrowhead). PrS, presubiculum; RS, retrosplenial cortex; S, subiculum. Anatomical delineations are based on (*73*). (**F**) Shoreline of Tanzania and Zanzibar in East Africa, with Latham Island. (**G**) Drone photo of Latham Island, together with all flight trajectories of bat 1. Different colors denote different sessions; green dots denote the two platform locations. Scale bar, 50 m. (**H**) Photos taken on the island, depicting some of the landmarks that the bats may have used for navigation. See also fig. S4A. [Photos: (Top) Shaked Palgi; (Bottom) Chen Cohen] (**I**) Photograph of a landing platform [the east green dot in (G)]. Platform height, 1.8 m. [Photo: Nachum Ulanovsky] (**J**) Session duration (top) and total flight time (bottom) for each of the 22 outdoor sessions from the six bats. (**K** to **O**) Histograms of behavioral data from flight epochs only, pooled over all the outdoor sessions. (K) Flight speed (low speeds indicate moments when the bat turned sharply) ($n = 208{,}278$ GPS samples). (L) Flight altitude. (Left) Altitude above mean sea level, based on the altimeter. (Right) Altitude above ground [the ground level was taken from a drone-based digital elevation map, as in fig. S3E, and was subtracted from the flight altitude (*42*)]. Maximal altitudes reached >20 m but were clipped here for display purposes. (M) Area covered by the bats in each experimental session ($n = 22$ sessions). (N) Percentage of the island that was visible to the bat during flight at each moment [$n = 8252$ time points (1-s intervals); see also fig. S6]. (O) Number of accumulated full rotations (360°) per session, calculated by integrating the absolute values of the bat's turning angles ($n = 22$ sessions). (**P**) Angular coverage per session, for bat 1 [each line represents one session and shows the total flight time in each directional bin; *y* axis clipped at 60 s for clarity (dotted line marks the clipping)]. Same sessions and colors as in (G). (**Q**) Angular coverage per bat, averaged across all the sessions of each bat (gray, $n = 6$ bats), and grand average (black). N, north; E, east; S, south; W, west. (**R**) Comparison of heading direction versus head direction. Shown is the distribution of heading direction (*y* axis) for each head direction (*x* axis), pooled over all 17 sessions with magnetometer data (*42*). Head direction and heading direction were highly correlated (circular-circular correlation: $\rho = 0.88$).



Movie 1. Drone video footage of Latham Island. Drone video of the island taken during both experimental seasons. This video highlights the visual richness of the island with many available landmarks; we propose that these landmarks were used by the bats to align their neural compass. This video footage was taken during the day, whereas the experiments were conducted at night—but we note that these bats are highly visual (*52*,*62*) and could therefore see these visual features even on overcast nights or when the Moon was below the horizon. The bats could also sense these landmarks using other sensory systems, such as echolocation. Finally, we note that the birds visible in this footage were asleep during the night hours when we conducted our experiments, and therefore they did not interfere with the experiments. [Video: Shaked Palgi]

$\rho = 0.88$); and because the measurements of heading direction were more reliable (*42*), we subsequently used the heading direction to construct directional tuning curves.

Head-direction cells exist outdoors and have similar characteristics to those recorded indoors

We first asked whether head-direction coding exists in bats flying outdoors and whether it shows similar properties to directional coding in indoor situations. We focused here on azimuthal directional tuning because neuronal tuning to pitch (vertical) direction could not be computed, as the bats exhibited a very restricted range of pitch angles in-flight (fig. S5C). A total of 406 well-isolated neurons passed our inclusion criteria based on behavioral coverage and minimum spikes [we moved our tetrodes daily to allow the recording of new neurons (*42*)]. Of these, 97 neurons (24%) were classified as significant head-direction cells on the basis of their unimodal directional tuning compared with spike shuffles [Fig. 2, A to C; for detailed criteria see (*42*)]. These 97 head-direction cells were the focus of the analyses below. The head-direction cells were recorded in the dorsal presubiculum (59/97) and in the retrosplenial cortex (38/97). We pooled together the data from both brain regions because we did not find major differences between the two (fig. S7). This percentage of head-direction cells is consistent with previous reports from laboratory-based studies of these brain areas in rats and bats (*15*, *21*, *22*, *25*, *43*) [but see (*44*, *45*)]. The directional tuning of these neurons was stable over time within the session (Fig. 2. A and B, bottom-right raster; Fig. 2D; median stability score: Pearson correlation coefficient $r = 0.81$). The cells exhibited strong unimodal tuning (fig. S8, A to C), with a typical tuning width of $117 \pm 35°$ (mean ± SD, tuning width at half-height; fig. S8D)—slightly wider than reported in the presubiculum of rodents (*15*, *46*). Head-direction cells remained stable when comparing flight and stationary nonflight periods (Fig. 2, E and F, and fig. S9), as reported for head-direction cells in laboratory setups (*22*, *25*).

The two most prominent directional cues available outdoors at night are the Moon and the wind. We did not find any overrepresentation of these two variables (Moon, fig. S8, E and F; wind, fig. S8, G to I). Rather, the preferred directions of head-direction cells outdoors covered uniformly all possible directions (Fig. 2, G and H; Kuiper's test for uniformity: $P = 0.67$), similar to studies indoors (*15*, *22*). We controlled for nonuniform behavioral coverage (Fig. 1, P and Q): There was no association between the behavioral-coverage distribution and the tuning curves of the neurons (fig. S8, J and K).

We also examined the neural code for head direction at the population level. First, we trained a Bayesian decoder to decode the bat's direction from the neural activity of simultaneously recorded neurons [$n = 10$ sessions with ≥10 cells, out of which ≥5 were significant head-direction cells; we used 10-fold cross-validation (*42*)]. The decoding was successful, with a median decoding error of 30° (Fig. 2, I and J), indicating that the population of presubiculum neurons carries reliable directional information. Second, we used the uniform manifold approximation and projection (UMAP) dimensionality-reduction algorithm to plot the low-dimensional neural manifold of the population activity of head-direction cells. Our analysis revealed a ring-shaped structure (Fig. 2K), consistent with recent findings in rodents and flies (*47*–*50*).

Overall, we conclude that head-direction cells exist outdoors and have basic characteristics similar to those of the "classical" head-direction cells recorded indoors. This allowed us to further test their coding properties outdoors.

Head-direction tuning remained stable over the geographical scale of the island, consistent with a global compass

We next addressed our second main question, the stability of the directional tuning over the geographical scale of the island. The bats never had a complete view of the island, owing to occlusions from cliffs, boulders, and tents (Fig. 1N and fig. S6), and hence the set of sensory cues available to them at each location was different. One hypothesis would therefore be that head-direction cells represent orientation in a local mosaic manner, with each neuron exhibiting different preferred directions at different locations (fig. S1, left). Another hypothesis is that head-direction cells might represent orientation in a global fashion, with each neuron showing the same preferred direction across space (fig. S1, right).

We first examined the stability of the head-direction tuning across space by splitting the data into two equal parts on the basis of location and calculating the head-direction tuning separately for each half (Fig. 3A). The head-direction tuning was highly stable between the two halves (Fig. 3A, bottom, and Fig. 3B, left). We also examined whether the island's shore, a prominent elongated landmark, anchors the head-direction tuning. This was tested by splitting the data on the basis of the bat's proximity to the west shore or the south shore, which are ~90° from each other (fig. S3D)—if the shoreline aligned the directional tuning, we would expect ~90° rotation of the head-direction tuning between these two areas. However, the directional tuning showed high similarity between the regions (Fig. 3, B to D: median differences of <30°, or <22% of the cells' tuning width). The stability of the tuning was also maintained across additional spatial bisections that we performed (figs. S10 and S11) as well as when we split the data according to other behavioral variables, such as altitude, speed, and angular velocity (fig. S12). This high stability was robust to various choices of thresholds (figs. S11, G and H, and S13, C to G) and was confirmed using a cross-decoding analysis (fig. S13B).

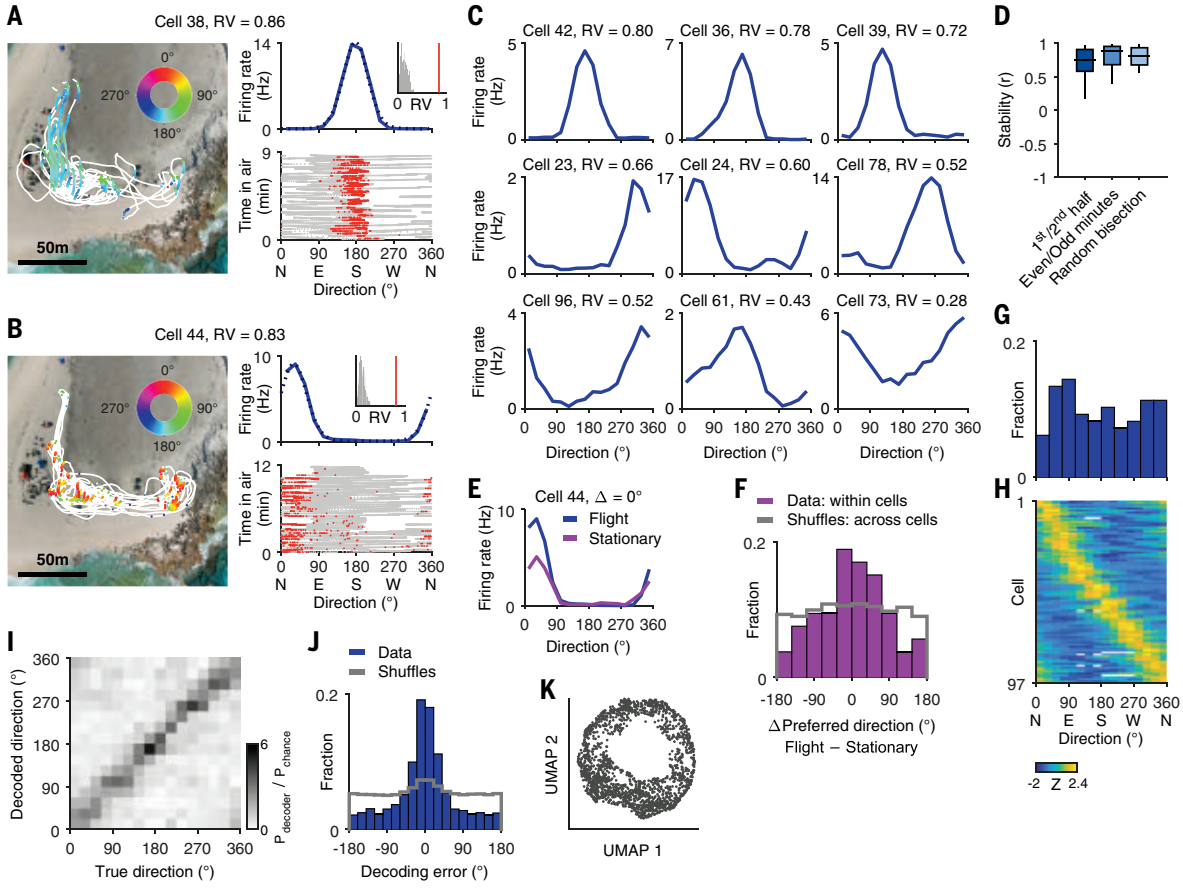


Fig. 2. Head-direction cells recorded outdoors. (**A**) Neuron with significant directional tuning. (Left) Flight trajectories (white) and spikes (colored dots), plotted on top of a drone photo of the island. Spike colors correspond to the bat's direction at the moment of firing; see circular color scale. (Top right) Directional tuning curve (blue), with fitted von Mises function (circular Gaussian; dotted line). (Bottom right) Bat's heading direction versus time (light gray), together with a spike raster (red dots). Time in air is the net flight time of the bat, from the session start to its end. (Inset) Rayleigh vector (RV) length of the cell's directional tuning curve (red line, RV = 0.86), and RV distribution for 1000 rigid spike-shuffles (gray). The RV of the data was larger than all 1000 shuffles. (**B**) Another head-direction cell (RV = 0.83); plotted as in (A). (**C**) Nine additional head-direction cells from five bats, showing their directional tuning curves. (**D**) Stability of directional tuning across the population of head-direction cells. The three boxplots show three different stability scores: Pearson correlation coefficient *r* for first half versus second half of the session ($n = 97$ cells), even versus odd minutes ($n = 82$ cells recorded for >2 min in air), and random bisections [$n = 97$ cells (*42*)]. Horizontal lines denote medians, boxes denote 25th–75th percentiles, and whiskers denote 10th–90th percentiles. (**E**) Head-direction tuning for a neuron recorded during flight (blue) and stationary nonflight (purple). Δ, absolute difference in preferred direction. (**F**) Histogram of ΔPreferred direction, calculated within cells (data: purple) and across cells (shuffles: gray). The differences in preferred direction were distributed around 0° (Rayleigh test for uniformity: $P = 0.0004$; $n = 53$ cells) (*42*). (**G**) Histogram of cells' preferred direction. Kuiper's test for uniformity: $P = 0.67$, indicating a uniform distribution. (**H**) Directional tuning for all the significant head-direction cells (rows, $n = 97$), sorted by the neuron's preferred direction. Tuning is *z*-scored (see color bar). (**I**) Decoding confusion matrix, pooled over the 10 sessions with enough neurons (*42*): the probability of decoded direction (*y* axis) for each true direction (*x* axis), normalized by the uniform chance probability $P_{chance} = 1/n_{bins}$ (see color bar). (**J**) Histogram of decoding errors (blue), pooled over the same 10 sessions as in (I). Median absolute error = 30°. Gray: decoding using rigid spike shuffles. Rayleigh test for uniformity: $P < 10^{-300}$, indicating highly nonuniform distribution; Wilcoxon rank-sum test comparing per-session median absolute error for data versus shuffle: $P = 0.002$. (**K**) UMAP dimensionality-reduction of the neuronal population activity in-flight, from a single experimental session, projected onto two UMAP components [using $n = 7$ simultaneously recorded head-direction cells; dots represent individual time points: 200-ms bins (*42*)], showing a ring-shaped neuronal manifold.

We next conducted a more systematic and fine-grained analysis of the head-direction tuning across space (Fig. 3, E to O). For each neuron, we binned the island into 40-by-40-m spatial bins (with a sliding-window step of 5 m) and computed the local directional tuning within each bin, generating two types of maps: (i) A preferred-direction map: the map of preferred directions across space (Fig. 3, G and J). (ii) A tuning correlation map: the map of Pearson correlations between the directional tuning computed within each bin and the tuning based on all the data outside that bin (Fig. 3, H and K). To quantify the degree of uniformity of these maps, we defined two indices of directional stability across space: (i) RV_{map}: the Rayleigh vector length of the distribution of preferred directions across space (computed from the preferred-direction maps; Fig. 3L) (*51*) and (ii) median correlation (computed from the tuning-correlation maps; Fig. 3M). We repeated these analyses using smaller bin sizes of 30-by-30 and 20-by-20 m (*42*). Overall, the neurons' preferred-direction maps were highly uniform (Fig. 3, G and J), corresponding to high values of RV_{map} (Fig. 3L, blue; median $RV_{map} = 0.66$ to 0.73 for the three spatial bin sizes)—very different from what is obtained using two types of shuffles [Fig. 3L, gray; here, we used both spike shuffles (dark gray) and cell shuffles (light gray), where we shuffled spatial bins across cells (*42*); 69 to 76% of the neurons were above the 95th percentile of both shuffles (significant cells)]. Similarly, the tuning correlation showed very high values across space (Fig. 3, H and K; Fig. 3M, blue; population median correlation: $r = 0.74$ to 0.75)—also very different from both types of shuffles (Fig. 3M, gray; 81 to 82% significant cells). The results were robust to the choice

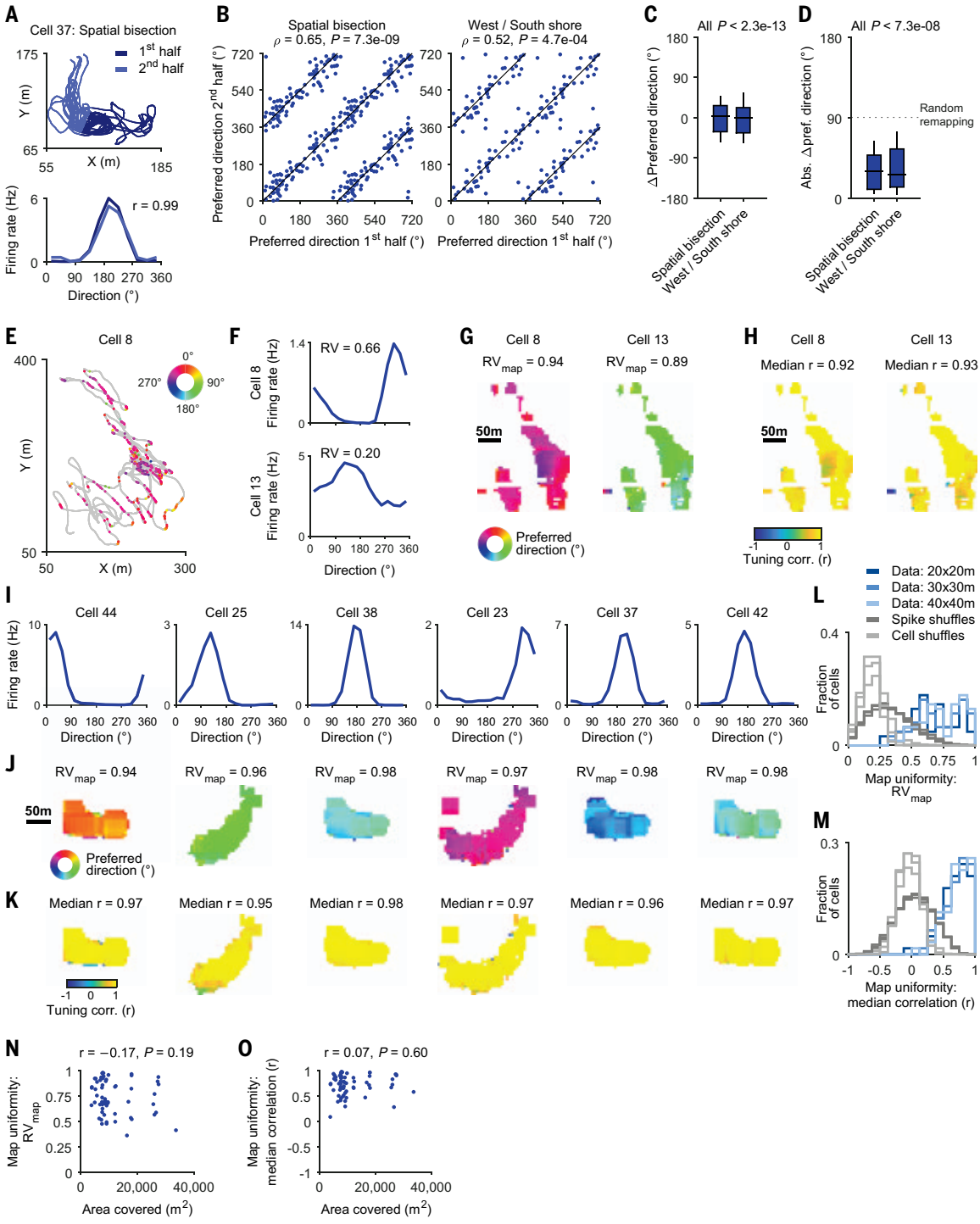


Fig. 3. Head-direction tuning remained stable across the geographical scale of the island. (**A**) Neuron showing stable tuning to spatial bisection of the bat's behavior. Trajectories (top) and directional tuning curves (bottom), plotted separately for the two halves of the data (light and dark blue). Note the high similarity of the two tuning curves (Pearson $r = 0.99$). (**B**) Preferred direction in the second half versus first half, for two types of spatial bisections: bisection along the long axis of the behavior (left) and separating the data on the basis of proximity to the west shore versus south shore (right) (*42*). Dots are neurons: $n = 85$ cells (left) and 50 cells (right) with enough data in each half. In each scatter, data are plotted twice from 0° to 360° and 360° to 720°, for display purposes; black lines, identity lines. Circular-circular correlation coefficients and P values are indicated. (**C**) Difference in preferred direction between the two halves (ΔPreferred direction; Rayleigh test for uniformity: All $P < 2.3 \times 10^{-13}$, indicating highly nonuniform distribution of differences). (**D**) Absolute value of ΔPreferred direction (Wilcoxon signed-rank test versus 90°: All $P < 7.3 \times 10^{-8}$). In all boxplots, horizontal lines denote medians, boxes denote 25th–75th percentiles, and whiskers denote 10th–90th percentiles. P values in (C) and (D) were Bonferroni-corrected for $n = 2$ comparisons; similar P values were obtained when comparing to cell shuffles. (**E** to **H**) Two simultaneously recorded head-direction cells. (E) Flight trajectories (gray curve) with overlayed spikes from one neuron (dots), colored by the bat's heading direction at the moment of firing (see circular color scale). (F) Directional tuning curves of the two neurons. RV: Rayleigh vector length of the tuning curve. (G) Preferred-direction maps. Each pixel's color represents the preferred direction of the directional tuning curve at that spatial bin. RV_{map}: Rayleigh vector length of the preferred-direction map (computed over all spatial bins). All maps were plotted using 40-m bins with 5-m steps (sliding window). (H) Tuning correlation

maps. Each pixel's color represents the Pearson correlation coefficient *r* between the directional tuning inside that spatial bin and the tuning computed using all the data outside this spatial bin. Median *r*: median tuning correlation across all spatial bins. (**I** to **K**) Six additional head-direction cells, showing (I) directional tuning curves, (J) preferred-direction maps, and (K) tuning correlation maps. Note the uniform preferred directions in (J) and high tuning correlations in (K), indicating high spatial stability of the head-direction tuning. (**L** and **M**) Population analysis quantifying the spatial stability of head-direction tuning curves using two different indices for map uniformity [computed for $n = 62$ cells with enough data (*42*)]. We repeated the calculation using three spatial bin sizes (20-by-20–, 30-by-30–, and 40-by-40–m bins; results plotted in different shades of blue) and using two types of shuffles: spike shuffles (dark gray) and cell shuffles (light gray). Note the higher values for the data compared with shuffles. (L) RV_{map}: Population medians = 0.66 to 0.73 (range across the three spatial bins); 69 to 76% significant cells compared with both shuffles. (M) Median tuning correlation: Population medians = 0.74 to 0.75; 81 to 82% significant cells. (**N**) No correlation between RV_{map} and the area covered by the bat [defined using the flight hull of the bat's spatial coverage (*42*); Pearson $r = -0.17$, $P = 0.19$; $n = 62$ cells]. (**O**) Same as (N), for the median tuning correlation versus area covered ($r = 0.07$, $P = 0.60$).

of spatial bin size (Fig. 3, L and M) and to the details of the shuffling procedure (fig. S14) and did not depend on the total area covered by the bats (Fig. 3, N and O).

Lastly, we explicitly searched for "mosaic cells," neurons that exhibit strong local head-direction tuning but only weak global tuning (figs. S1, left, and S15, A to H). To this end, we calculated a third type of map, where for each spatial bin we quantified the sharpness of the local directional tuning through its Rayleigh vector length (fig. S15D). We identified neurons with significant local tuning using shuffling, as before (*42*). This procedure yielded 32/213 (15%) putative mosaic neurons (fig. S15E). Notably, these putative mosaic neurons did not exhibit sharp local tuning (fig. S15F, red dots)—their local tuning was much less sharp than the local tuning of head-direction cells (fig. S15H). In fact, we found no compelling evidence for true mosaic neurons, as there were almost no neurons with strong local but weak global tuning (fig. S15F: no neurons in the top-left corner; and see fig. S15, A to D, for some of the most sharply tuned mosaic cells). Thus, these cells are possibly a separate weakly tuned functional class of neurons, rather than head-direction cells that remap across space.

Together, these analyses suggest that head-direction cells recorded outdoors encode orientation in a global manner over a large geographical region, which supports the global compass hypothesis (fig. S1, right) and argues against the local mosaic hypothesis (fig. S1, left).

Head-direction tuning remained stable despite dynamics of the Moon

We next addressed our third main question of whether the head-direction neural compass remains stable despite the movement of celestial cues. Because Egyptian fruit bats are nocturnal and highly visual animals (*4*, *52*–*54*), we sought to determine whether their neural compass anchors to the Moon.

In some of the sessions, we recorded head-direction cells before moonrise, waited for a few hours, and then recorded the same neurons again when the Moon had become visible (Fig. 4, A to C). The directional tuning of head-direction cells was highly stable when comparing flights with or without the Moon (Fig. 4, B and C; population: Fig. 4, D to G, light blue).

Stars and the Milky Way are also often visible in natural environments (*55*), and these cues were shown to aid animal navigation (*41*, *56*, *57*). Therefore, we compared the directional tuning of neurons when the Moon and stars were visible against periods of cloud coverage (which transiently obscured both the Moon and the stars). Again, the directional tuning of head-direction cells was highly stable when comparing flights with Moon and stars visible versus cloud coverage (Fig. 4, D to G, dark blue).

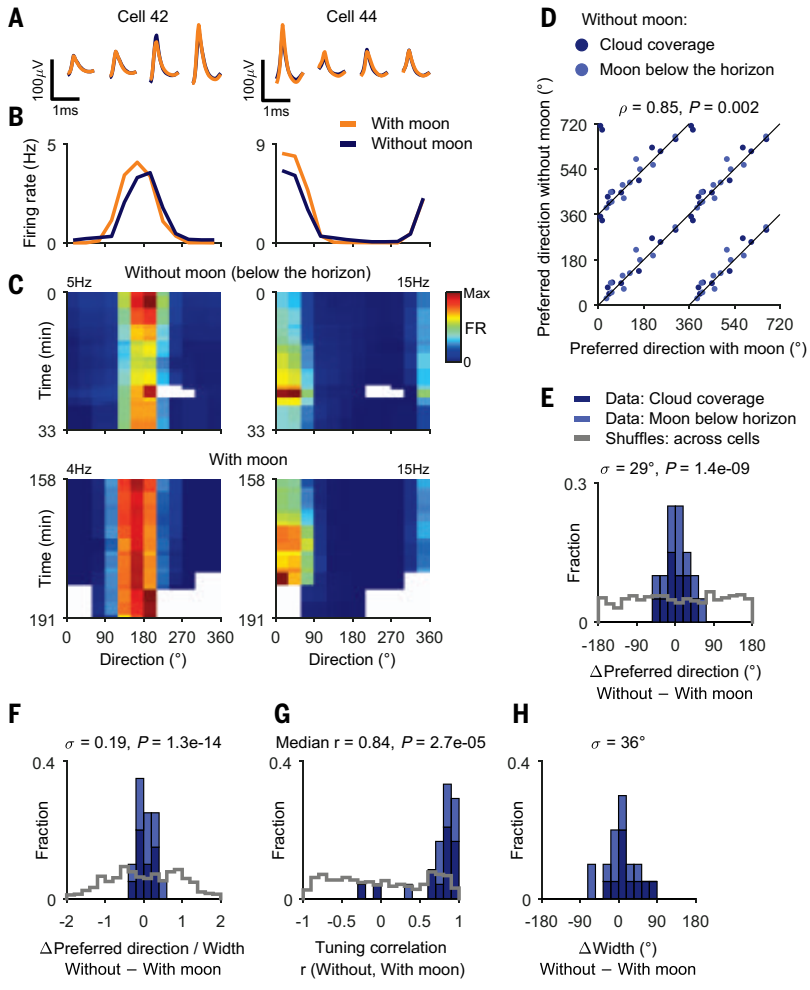


Fig. 4. Head-direction tuning remained stable despite dynamics of celestial cues. (**A** to **C**) Two neurons recorded simultaneously on different tetrodes. (A) Mean spike waveforms, plotted separately for the two parts of the recording session (dark blue: without Moon, orange: with Moon; the four lines correspond to the four channels of the tetrode). (B) Directional tuning curves, computed separately for the two parts of the session [same colors as in (A)]. (C) Firing rate (FR) as a function of direction (*x* axis) and time (*y* axis), for the two parts of the session: without Moon (top) and with Moon (bottom). White indicates bins without sufficient behavioral data. (**D** to **H**) Population analyses of cells recorded with and without Moon. In total, $n = 20$ neurons from four bats had enough data in both conditions (*42*); "Without moon" consists of cases when the Moon was below the horizon (light blue; $n = 10$ cells) or when the Moon was behind cloud coverage (dark blue; $n = 10$ cells). (D) Preferred direction without Moon versus preferred direction with Moon. Data are plotted twice from 0° to 360° and 360° to 720°, for display purposes; black lines are the identity lines; circular-circular correlation: $\rho = 0.85$, $P = 0.002$. (E) Stacked histograms of ΔPreferred direction (°), without Moon minus with Moon (blue), and the population cell shuffles (gray); σ, standard deviation. Rayleigh test for uniformity: $P = 1.4 \times 10^{-9}$. (F) Same as (E), with each cell normalized by its overall tuning width. Rayleigh test for uniformity: $P = 1.3 \times 10^{-14}$. (G) Histogram of tuning correlation (Pearson *r*) of the tuning curves with versus without Moon. Population median correlation: $r = 0.84$. Wilcoxon signed-rank test versus 0: $P = 2.7 \times 10^{-5}$. (H) Histogram of ΔWidth (°), Without–With Moon. Wilcoxon signed-rank test versus 0°: $P = 0.30$.

Thus, we observed exceptional stability of the head-direction tuning under both types of comparisons: (i) Moon visible versus Moon below the horizon, and (ii) Moon and stars visible versus cloud coverage. In both cases, we did not observe changes in the head-direction tuning—as manifested by the stability of the neurons' preferred direction (Fig. 4D, circular-circular correlation: $\rho = 0.85$, $P = 0.002$; Fig. 4, E and F, Rayleigh test for uniformity: $P = 1.4 \times 10^{-9}$ and $P = 1.3 \times 10^{-14}$), their high tuning correlations (Fig. 4G; median correlation: $r = 0.84$; Wilcoxon signed-rank test versus 0: $P = 2.7 \times 10^{-5}$), and their stable tuning width (Fig. 4H; Wilcoxon signed-rank test versus 0°: $P = 0.30$). A decoding analysis further confirmed that the Moon's presence does not contribute to the directional coding of head-direction cells (fig. S16, B to D).

Head-direction tuning stabilized over several nights of exploring a novel real-world environment

This was the first-ever exposure of these bats to this environment, which presented an opportunity to study the ontogeny of the head-direction signal in a new natural environment. We examined the dynamics of the head-direction tuning across multiple nights, in search of signatures of learning. Because our neural-recording method does not allow tracking the same individual neurons over consecutive nights, we examined whether there are population-level differences in the stability of the head-direction tuning between neurons that were recorded during the first nights outdoors and neurons recorded on later nights. Although the cells were generally very stable (Figs. 2D and 3M), we hypothesized that cells recorded in earlier nights might be less stable.

For each neuron, we examined the prevalence of rapid within-session dynamics. We used the data from the first 10 min of the session and compared the tuning in the first 5 min (minutes 0 to 5) with the tuning in the next 5 min (minutes 5 to 10). Some of the neurons recorded over the first few nights exhibited large changes in the first 10 min of the session—changes that were not observed in neurons recorded on later nights (Fig. 5, A and B; population: Fig. 5C, left). Similar results were found when repeating this calculation using the last 10 min of the sessions (Fig. 5C, right). Thus, we observed transient dynamics of the head-direction tuning in the first few nights of flight outdoors, which slowly stabilized as the bats explored the island on consecutive nights [Fig. 5C; Spearman correlations: session start (first 10 min): $\rho = -0.38$, $P = 0.002$; session end (last 10 min): $\rho = -0.50$, $P = 0.0005$]. We also calculated the head-direction tuning along the session in time windows of 15 min and fitted a circular-linear regression to the preferred direction of each neuron as a function of time within the session (*42*). We observed faster drift of the preferred direction (steeper regression slopes) in early nights than in later nights (Fig. 5D; $\rho = -0.60$, $P = 10^{-8}$), which is consistent with a gradual stabilization of the directional code. The directional coding reached high stability on later nights, as indicated by very small changes of the preferred direction during the session (Fig. 5D, nights 3 to 6).

To examine potential changes in the spatial stability of directional tuning across nights, we tested whether neurons' median correlation index (as in Fig. 3, K and M) varied as a function of their recording night. Neurons became more spatially stable as the bats flew more nights outdoors (Fig. 5E; $\rho = 0.31$, $P = 0.015$).

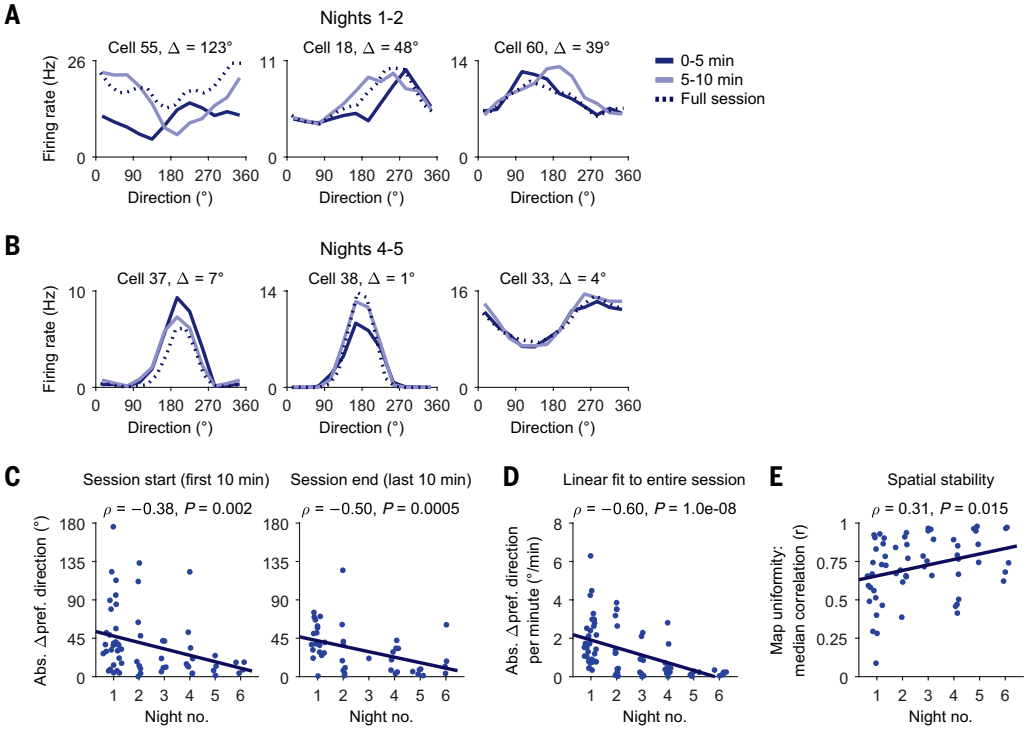


Fig. 5. Head-direction tuning stabilized across nights. (**A** and **B**) Neurons recorded on different nights. (A) Three neurons recorded on nights 1 and 2. Tuning was calculated separately for the first 5 min of the session (dark blue), for the next 5 min (light blue), and for the overall session (dashed lines). The neuronal tuning changed between these 5-min segments; Δ indicates the absolute ΔPreferred direction. (B) Three different neurons, recorded on nights 4 and 5; plotted as in (A). (**C**) Directional-tuning changes across nights. (Left) Absolute difference between the preferred direction calculated during the first 5 min and the next 5 min, plotted as a function of night of recording ($n = 62$ neurons with enough data; dark blue line, regression line; Spearman correlation: $\rho = -0.38$, $P = 0.002$). (Right) As in left panel, but comparing the last 5 min to the penultimate 5 min ($n = 45$ neurons; Spearman correlation: $\rho = -0.50$, $P = 0.0005$). (**D**) Slopes of within-session linear fits to the preferred direction of each neuron; the slopes are plotted here against night number, as in (C) ($n = 77$ neurons with enough data; Spearman correlation: $\rho = -0.60$, $P = 10^{-8}$). (**E**) Spatial stability across nights. Median tuning correlation (same index plotted in Fig. 3M) as a function of the night of recording ($n = 60$ neurons with enough data; Spearman correlation: $\rho = 0.31$, $P = 0.015$); plotted as in (C). The dots in (C) to (E) were slightly jittered horizontally, for display purposes only.

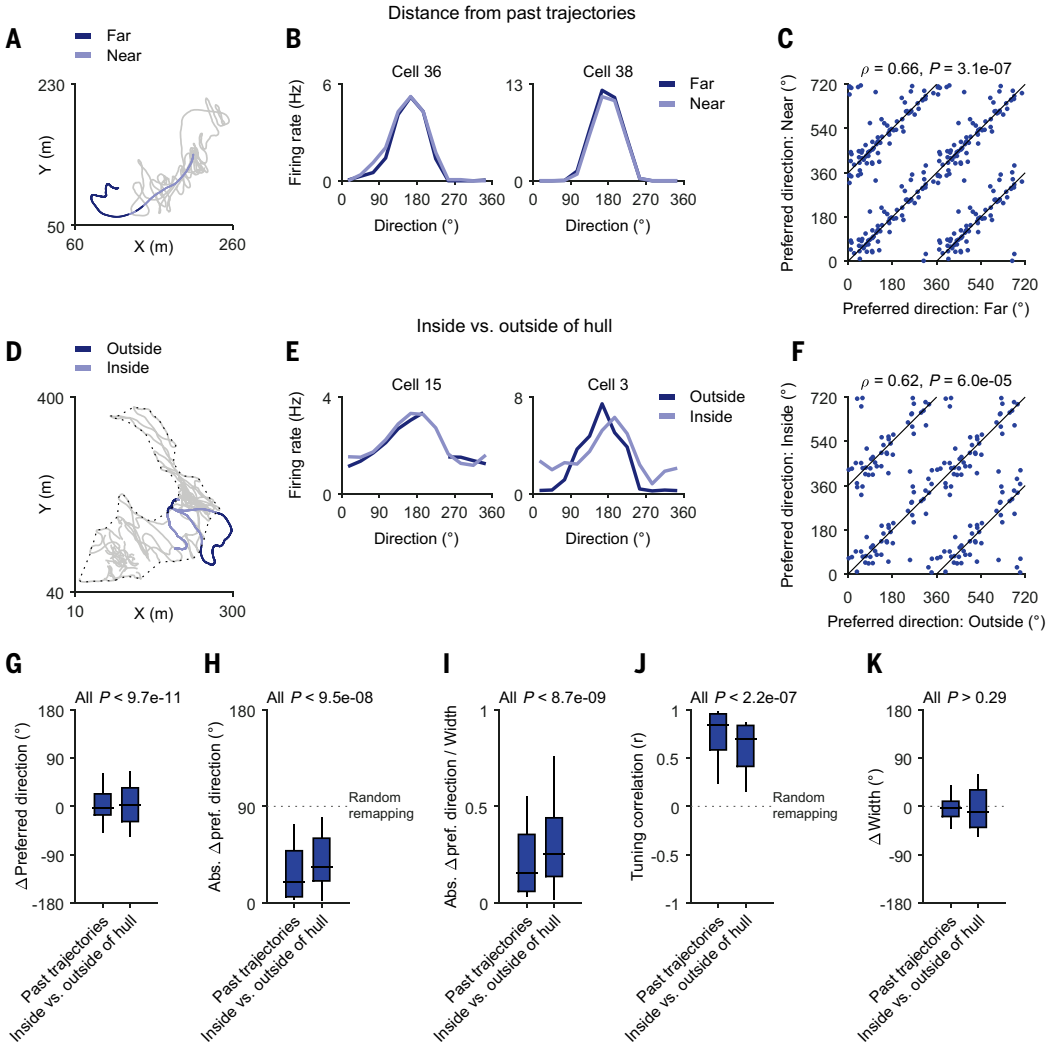


Fig. 6. Head-direction tuning remained stable during novel trajectories. (**A** to **C**) Bisection of the data by the distance to the past trajectories of the bat in the same session. (A) Example flight trajectory, colored by the distance to the past flight trajectories in this session (light blue, near; dark blue, far). Gray trace represents all the past trajectories in this session (the set of past trajectories was dynamically updated on a flight-by-flight basis). The "far" portion of this flight had a duration of 21 s. (B) Two neurons with enough data for this bisection (recorded simultaneously on different tetrodes). For each cell, two directional tuning curves are plotted, computed from all the flights in the two halves of the data [same colors as in (A)]. (C) Population scatterplot of the preferred direction in the second half (near) versus first half (far), plotted as in Fig. 3B (circular-circular correlation: $\rho = 0.66$, $P = 3.1 \times 10^{-7}$; $n = 76$ head-direction cells with enough data in both halves). (**D** to **F**) Split of the data by the dynamic flight hull, which is based on the entire history of the bat across all sessions. (D) Example flight, plotted as in (A) (light blue, inside the flight hull; dark blue, outside the flight hull; gray, past flights in all the previous sessions; dashed line, dynamic flight hull). The "outside hull" portion of this flight had a duration of 36 s. (E) Two neurons, with the two colors indicating the two splits of the data [same colors as in (D)]. (F) Population scatterplot for this bisection, plotted as in (C) ($\rho = 0.62$, $P = 6 \times 10^{-5}$; $n = 47$ cells). (**G** to **K**) Population boxplots for both types of bisections. Horizontal lines in boxplots denote medians, boxes denote 25th–75th percentiles, whiskers denote 10th–90th percentiles. (G) Difference in preferred direction between the two halves (ΔPreferred direction; Rayleigh test for uniformity: all $P < 9.7 \times 10^{-11}$). (H) Absolute value of ΔPreferred direction (Wilcoxon signed-rank test versus 90°: all $P < 9.5 \times 10^{-8}$). (I) Absolute value of ΔPreferred direction, normalized by each cell's overall tuning width (Kolmogorov-Smirnov test versus cell shuffles: all $P < 8.7 \times 10^{-9}$. (J) Tuning correlation (Pearson r) between the first half and second half (Wilcoxon signed-rank test versus 0: all $P < 2.2 \times 10^{-7}$). (K) Differences in tuning width between the first half and second half (Wilcoxon signed-rank test versus 0°: all $P > 0.29$). All P values in (G) to (K) were Bonferroni-corrected for the $n = 2$ comparisons; we note that similar P values were obtained when comparing with cell shuffles.

These dynamics of the head-direction code across nights could not be explained by changes in behavior (fig. S17, P to U), changes in spike-sorting quality (fig. S17, J to O), or differences in anatomical recording areas (fig. S17, A to I).

Head-direction tuning remained stable during flights outside a bat's familiar area

Is the head-direction tuning stable when a bat flies for the first time over a new region of the island? We first bisected the data on the basis of its spatial distance from all the previous flights in the same session (Fig. 6A) (*42*). This analysis showed stable directional tuning when comparing moments when the bat was near versus far away from previous flights of the same session (Fig. 6, B and C; near versus far: $\rho = 0.66$, $P = 3.1 \times 10^{-7}$). We then defined for each bat a behavioral "flight hull," which encompassed the region of the island that this bat had already explored during this session and previous sessions (*42*). This allowed us to compare moments when the bat flew over a region of the island that it had already explored ("inside hull") against flights over regions it had never visited before ("outside hull"; Fig. 6D, example), which again showed stable directional tuning for familiar versus novel trajectories (Fig. 6, E and F; inside versus outside hull: $\rho = 0.62$, $P = 6.0 \times 10^{-5}$). Both types of analyses showed high stability of the

directional tuning compared with random remapping (random rotation) of the tuning curve (Fig. 6, G to J), with no significant changes in tuning width (Fig. 6K; see also fig. S18, A and B, for varying distance thresholds from past trajectories and from the flight hull).

Even though these bats generally avoid prolonged flights over water, we observed two cases where bats flew far away from the island, toward the ocean (fig. S18C; see also fig. S2B, bats 3 and 6; flight durations = 38 and 104 s, respectively; maximal distance from shore = 118 and 282 m, respectively). Our fast boat waited in the water at night and safely captured these bats in both cases. Thus, we could compare the head-direction tuning above land to the tuning above the ocean. Although only a small number of head-direction cells were recorded during these two extreme flights, these few neurons showed stability of their directional tuning over land and over the ocean (fig. S18D).

Discussion

Here we have reported single-unit neural recordings conducted on a remote island where fruit bats navigated freely (Fig. 1). We identified head-direction cells that showed similar properties to classical head-direction cells (Fig. 2), supporting the relevance of head-direction cells for outdoor navigation. These neurons exhibited highly stable directional tuning across the island's geography (Fig. 3), suggesting a global representation of direction in the head-direction network (fig. S1, right), rather than a mosaic representation (fig. S1, left), at least on the scale of the island. The stable head-direction tuning was also robust to the dynamics of celestial cues, such as the movement of the Moon and stars (Fig. 4). Together, these results suggest that head-direction cells can serve as a stable and reliable neural compass for real-world navigation outdoors.

A potential limitation of this study is that we computed tuning curves to heading direction rather than head direction, because heading-direction measurements (Fig. 1C) were more precise and more technically reliable in our data. We note that head direction and heading direction are highly correlated in Egyptian fruit bats (Fig. 1R, $\rho = 0.88$). However, they can sometimes differ, potentially yielding better coding for head direction during these rare moments of mismatch (*58*). This subtle difference could potentially explain why we observed slightly wider tuning during high-flight-curvature moments (fig. S12F), which are the only moments when head direction and heading direction are expected to differ (*59*).

What are the sensory cues that the bats used to generate their stable neuronal sense of direction? For short timescales of several seconds, the bats may have used internal sensory cues via path-integration mechanisms. However, path integration accumulates errors over the course of multiple turns (*60*, *61*), and it is therefore unreliable over longer timescales of minutes. As the bats performed dozens of angular turns per session (Fig. 1O), they must have also used external cues to orient themselves. Egyptian fruit bats have excellent olfaction, hearing, and vision (*53*, *62*), so potentially all these sensory modalities can be used to anchor their compass. There are two categories of sensory cues that can be used for directional orientation: (i) Global directional cues, whose angles relative to the animal do not depend on its position, allowing for easy computation of head direction; these cues include the direction of Earth's magnetic field, the direction of the wind, the sound of breaking waves, and the angle of celestial cues. (ii) Local landmarks (including the geometry of the environment), perceived through vision, olfaction, or echolocation. It is likely that, on the island, the bats used local landmarks more than global cues, for the following reasons.

1) Magnetoreception: The use of magnetoreception by our bats is inconsistent with the dynamics observed in the neural code across nights (Fig. 5), because Earth's magnetic field is constantly present and therefore was available to the bats from the first moment. It is also inconsistent with results of additional experiments that we conducted inside a flight tent on the island: We found that the head-direction tuning rotated between indoor flight and outdoor flight (fig. S19), which is inconsistent with magnetic alignment, because Earth's magnetic field is identical indoors and outdoors. The use of magnetoreception is also unlikely given a separate laboratory magnetic-manipulation experiment that we conducted: See supplementary results in the supplementary materials and fig. S20. Finally, we note that there is no compelling evidence to date for magnetoreception in any nonmigratory bat species, Egyptian fruit bats included.

2) Olfaction and wind direction: For olfaction to be a reliable global directional cue, the wind direction needs to be stable, which is not the case in most natural environments, including in our recordings (fig. S8G). Furthermore, we could not find any correlation between the variability in wind direction and the variability in the neuronal preferred direction (fig. S8I). This argues against the notion that the head-direction tuning was anchored by smelling wind-carried odorants or by sensing the wind direction itself (through somatosensation).

3) Hearing: The sounds of the breaking sea waves are unlikely to be a useful global cue for bats, as waves can appear from multiple directions and also because most of the acoustic energy of breaking waves is at low frequencies (*63*), below the minimum frequency that these bats can hear (*64*). We also did not find stable directional tuning inside the tent compared with outdoors (fig. S19, D and E), although the sound of waves was available in both conditions.

4) Vision: The main global visual directional cues are distant celestial cues (Moon, stars) or gradients of light across the sky (from the Moon below the horizon, or from distant human activity). However, their sole use was unlikely, because head-direction cells exhibited almost identical directional tuning both with and without the presence of the Moon and also during overcast nights with cloud coverage, when none of the distant celestial or anthropogenic visual cues were available (Fig. 4, B to G). Nevertheless, it is possible that when the Moon and stars are visible, bats are able to maintain a sense of direction by compensating for the dynamics of these celestial cues, similar to the ability of some species to compensate for movements of the Sun (*2*, *35*, *36*). However, this hypothesis seems unlikely because the dynamics of the Moon are much harder to predict and compensate for than the dynamics of the Sun (*40*) and because the moonrise–moonset cycle is decorrelated from nighttime. Nevertheless, it may be that when celestial cues and wind were available, the bats used them to transiently anchor their compass.

This leaves us with the second main possibility, namely, that the head-direction tuning is computed by learning a configuration of local landmarks (including the island's geometrical features). Landmark-based navigation is a viable strategy outdoors because natural environments are highly structured and contain extremely rich landmark information, across modalities (*5*, *65*) (see Fig. 1H and fig. S4 for the high visual richness of the island). These landmarks can be perceived by any available modality, such as vision, olfaction, or echolocation, or through their multisensory combination (*5*, *52*, *62*, *66*, *67*). We believe that vision is the dominant modality for sensing landmarks in our experiments, because Egyptian fruit bats have excellent vision (*4*, *52*, *53*, *62*, *68*) and because vision has a longer range than the other senses. We propose that using local landmarks is the main strategy used by bats in real-world, cue-rich environments.

Regardless of the sensory modality, bats need to learn a set of multiple landmarks. They cannot rely on a single landmark, for two reasons: (i) In such large environments, not all landmarks are visible from all locations (Fig. 1N and fig. S6). (ii) If a single landmark had been used, the angle to it would change continuously as the bat moves through space, creating a "pinwheel" representation, wherein the preferred direction rotates around the focal landmark. However, this is not what we found; instead, we discovered a highly stable representation of direction throughout the space of the island (Fig. 3). This implies that a non-trivial computation must be performed to combine information from multiple landmarks to extract the bat's direction. Such computations were investigated previously in several theoretical studies, which suggested algorithmic solutions for generating stable head-direction coding from a set of landmarks (*69*, *70*). This computation was proposed to involve a number of brain regions, including the hippocampus, presubiculum, and retrosplenial cortex (*69*). As this computation relies on learning the

spatial arrangement of multiple landmarks, this learning process would need some time to take place and to stabilize in a completely new environment such as the island in our study (*67*, *69*, *71*). Indeed, we found here that the head-direction representation was less stable in the first two nights of exposure to the island and stabilized over the course of a few nights (Fig. 5). Notably, these models also allow the network to ignore unreliable dynamic cues [see also (*71*, *72*)]. This might explain why in our outdoor experiments, the head-direction cells seemed to ignore the easier-to-compute, but unreliable, dynamic celestial cues. Instead, the cells relied on the more computationally demanding but highly reliable information provided by multiple static landmarks.

Another possibility is that bats use a combination of global cues and local landmarks, wherein celestial cues participate in the initial binding of the set of landmarks into a global compass. For five of the six bats, the first night outdoors was a night with visible Moon and stars. We therefore hypothesize that on the first night, these bats could potentially anchor their compass using the clearly visible celestial cues, such as the Moon and Milky Way. This initial anchoring could accelerate the learning process by providing ground-truth directional information, and it can also explain the existence of a subpopulation of neurons already stably tuned on the first night (Fig. 5, C to E). Subsequently, on later nights, after the bats learned the landmark configuration on the island, their neural compass became independent of the celestial cues, and then they could use the landmarks to navigate on moonless nights (Fig. 4).

Overall, the results of this study imply that head-direction cells can indeed serve as a neural compass. Our study emphasizes the need to investigate the neurobiology of navigation in the wild. More generally, we call for broadening the scope of brain research toward neuroscience in the real world.

Materials and methods summary

We conducted wireless tetrode recordings in the dorsal presubiculum and retrosplenial cortex of bats that flew unconstrained over a remote oceanic island near East Africa. We caught adult Egyptian fruit bats (*R. aegyptiacus*) on the mainland, implanted them with a 16-tetrode microdrive, and then sailed to the island, where we conducted nightly recordings, one bat at a time. On the island, we kept moving the tetrodes to allow the recording of different neurons every night. We recorded single-unit neuronal activity together with behavioral measurements collected by onboard sensors: precise localization and flight direction (using GPS), kinematics (using an accelerometer and magnetometer), and altitude (using a pressure-based altimeter). We used a drone to obtain a 10-cm-resolution elevation map of the island and calibrated it with the GPS and altimeter. We also measured the wind direction using a portable weather station and logged Moon visibility online; Moon direction was calculated offline from astronomical databases. We identified head-direction cells on the basis of the Rayleigh vector length of their tuning curves compared with shuffles, and we tested their spatial stability by calculating the local directional tuning across the island. On some nights, we recorded both when the Moon and stars were visible in the sky and when they were occluded by heavy clouds or when the Moon was below the horizon, allowing us to compare these conditions. For some of the bats, we also recorded the same neurons inside a flight tent, comparing directional tuning indoors and outdoors. Finally, we conducted an experiment in our flight tunnel in Israel, where we recorded neurons in dorsal hippocampus area CA1 (cornu ammonis 1) using similar methods while transiently manipulating the magnetic field inside the tunnel. Further details are found in the supplementary materials and methods (*42*).

REFERENCES AND NOTES

1. R. Kays, M. C. Crofoot, W. Jetz, M. Wikelski, Terrestrial animal tracking as an eye on life and planet. *Science* **348**, aaa2478 (2015). doi: 10.1126/science.aaa2478; pmid: 26068858
2. H. G. Wallraff, *Avian Navigation: Pigeon Homing as a Paradigm* (Springer, 2005).
3. R. Wehner, *Desert Navigator: The Journey of an Ant* (Harvard Univ. Press, 2020).
4. A. Tsoar *et al.*, Large-scale navigational map in a mammal. *Proc. Natl. Acad. Sci. U.S.A.* **108**, E718–E724 (2011). doi: 10.1073/pnas.1107365108; pmid: 21844350
5. A. Goldshtein *et al.*, Acoustic cognitive map-based navigation in echolocating bats. *Science* **386**, 561–567 (2024). doi: 10.1126/science.adn6269; pmid: 39480949
6. O. Lindecke, A. Elksne, R. A. Holland, G. Pētersons, C. C. Voigt, Experienced migratory bats integrate the sun's position at dusk for navigation at night. *Curr. Biol.* **29**, 1369–1373.e3 (2019). doi: 10.1016/j.cub.2019.03.002; pmid: 30955934
7. S. Engels *et al.*, Anthropogenic electromagnetic noise disrupts magnetic compass orientation in a migratory bird. *Nature* **509**, 353–356 (2014). doi: 10.1038/nature13290; pmid: 24805233
8. W. Wiltschko, R. Wiltschko, Magnetic compass of European robins. *Science* **176**, 62–64 (1972). doi: 10.1126/science.176.4030.62; pmid: 17784420
9. M. Zapka *et al.*, Visual but not trigeminal mediation of magnetic compass information in a migratory bird. *Nature* **461**, 1274–1277 (2009). doi: 10.1038/nature08528; pmid: 19865170
10. J. O'Keefe, J. Dostrovsky, The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* **34**, 171–175 (1971). doi: 10.1016/0006-8993(71)90358-1; pmid: 5124915
11. J. O'Keefe, Place units in the hippocampus of the freely moving rat. *Exp. Neurol.* **51**, 78–109 (1976). doi: 10.1016/0014-4886(76)90055-8; pmid: 1261644
12. M. Fyhn, S. Molden, M. P. Witter, E. I. Moser, M.-B. Moser, Spatial representation in the entorhinal cortex. *Science* **305**, 1258–1264 (2004). doi: 10.1126/science.1099901; pmid: 15333832
13. T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, E. I. Moser, Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**, 801–806 (2005). doi: 10.1038/nature03721; pmid: 15965463
14. J. B. Ranck Jr., "Head direction cells in the deep cell layer of dorsal presubiculum in freely moving rats" in *Electrical Activity of the Archicortex*, G. Buzsáki, C. H. Vanderwolf, Eds. (Akademiai Kiado, 1985), pp. 217–220.
15. J. S. Taube, R. U. Muller, J. B. Ranck Jr., Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *J. Neurosci.* **10**, 420–435 (1990). doi: 10.1523/JNEUROSCI.10-02-00420.1990; pmid: 2303851
16. J. S. Taube, R. U. Muller, J. B. Ranck Jr., Head-direction cells recorded from the postsubiculum in freely moving rats. II. Effects of environmental manipulations. *J. Neurosci.* **10**, 436–447 (1990). doi: 10.1523/JNEUROSCI.10-02-00436.1990; pmid: 2303852
17. D. Yoganarasimha, X. Yu, J. J. Knierim, Head direction cell representations maintain internal coherence during conflicting proximal and distal cue rotations: Comparison with hippocampal place cells. *J. Neurosci.* **26**, 622–631 (2006). doi: 10.1523/JNEUROSCI.3885-05.2006; pmid: 16407560
18. M. P. Brandon, A. R. Bogaard, N. W. Schultheiss, M. E. Hasselmo, Segregation of cortical head direction cell assemblies on alternating theta cycles. *Nat. Neurosci.* **16**, 739–748 (2013). doi: 10.1038/nn.3383; pmid: 23603709
19. L. M. Giocomo *et al.*, Topography of head direction cells in medial entorhinal cortex. *Curr. Biol.* **24**, 252–262 (2014). doi: 10.1016/j.cub.2013.12.002; pmid: 24440398
20. A. Peyrache, M. M. Lacroix, P. C. Petersen, G. Buzsáki, Internally organized mechanisms of the head direction sense. *Nat. Neurosci.* **18**, 569–575 (2015). doi: 10.1038/nn.3968; pmid: 25730672
21. D. E. Angelaki *et al.*, A gravity-based three-dimensional compass in the mouse brain. *Nat. Commun.* **11**, 1855 (2020). doi: 10.1038/s41467-020-15566-5; pmid: 32296057
22. A. Finkelstein *et al.*, Three-dimensional head-direction coding in the bat brain. *Nature* **517**, 159–164 (2015). doi: 10.1038/nature14031; pmid: 25470055
23. N. Ulanovsky, *Natural Neuroscience: Toward a Systems Neuroscience of Natural Behaviors* (MIT Press, 2025). doi: 10.7551/mitpress/10813.001.0001
24. T. Wolbers, J. M. Wiener, Challenges for identifying the neural mechanisms that support spatial navigation: The impact of spatial scale. *Front. Hum. Neurosci.* **8**, 571 (2014). doi: 10.3389/fnhum.2014.00571; pmid: 25140139
25. J. S. Taube, The head direction signal: Origins and sensory-motor integration. *Annu. Rev. Neurosci.* **30**, 181–207 (2007). doi: 10.1146/annurev.neuro.29.051605.112854; pmid: 17341158
26. E. I. Moser, M.-B. Moser, B. L. McNaughton, Spatial representation in the hippocampal formation: A history. *Nat. Neurosci.* **20**, 1448–1464 (2017). doi: 10.1038/nn.4653; pmid: 29073644
27. M. B. Zugaro, A. Berthoz, S. I. Wiener, Background, but not foreground, spatial cues are taken as references for head direction responses by rat anterodorsal thalamus neurons. *J. Neurosci.* **21**, RC154 (2001). doi: 10.1523/JNEUROSCI.21-14-j0001.2001; pmid: 11425881
28. J. P. Goodridge, J. S. Taube, Preferential use of the landmark navigational system by head direction cells in rats. *Behav. Neurosci.* **109**, 49–61 (1995). doi: 10.1037/0735-7044.109.1.49; pmid: 7734080
29. R. W. Stackman, E. J. Golob, J. P. Bassett, J. S. Taube, Passive transport disrupts directional path integration by rat head direction cells. *J. Neurophysiol.* **90**, 2862–2874 (2003). doi: 10.1152/jn.00346.2003; pmid: 12890795
30. P. A. Dudchenko, L. E. Zinyuk, The formation of cognitive maps of adjacent environments: Evidence from the head direction cell system. *Behav. Neurosci.* **119**, 1511–1523 (2005). doi: 10.1037/0735-7044.119.6.1511; pmid: 16420155
31. J. S. Taube, H. L. Burton, Head direction cell activity monitored in a novel environment and during a cue conflict situation. *J. Neurophysiol.* **74**, 1953–1971 (1995). doi: 10.1152/jn.1995.74.5.1953; pmid: 8592189
32. R. M. Yoder *et al.*, Both visual and idiothetic cues contribute to head direction cell stability during navigation along complex routes. *J. Neurophysiol.* **105**, 2989–3001 (2011). doi: 10.1152/jn.01041.2010; pmid: 21451060

33. A. E. Smith, E. R. Wood, P. A. Dudchenko, The stimulus control of local enclosures and barriers over head direction and place cell spatial firing. *Brain Behav.* **11**, e02070 (2021). doi: 10.1002/brb3.2070; pmid: 33606361
34. K. J. Jeffery, The mosaic structure of the mammalian cognitive map. *Learn. Behav.* **52**, 19–34 (2024). doi: 10.3758/s13420-023-00618-9; pmid: 38231426
35. T. W. Cronin, S. Johnsen, N. J. Marshall, E. J. Warrant, "Visual orientation and navigation" in *Visual Ecology* (Princeton Univ. Press, 2014), pp. 289–312.
36. J. L. Gould, Sensory bases of navigation. *Curr. Biol.* **8**, R731–R738 (1998). doi: 10.1016/S0960-9822(98)70461-0; pmid: 9778524
37. A. Rozhok, *Orientation and Navigation in Vertebrates* (Springer, 2008).
38. B. el Jundi *et al.*, Neural coding underlying the cue preference for celestial orientation. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11395–11400 (2015). doi: 10.1073/pnas.1501272112; pmid: 26305929
39. H. Mouritsen, Long-distance navigation and magnetoreception in migratory animals. *Nature* **558**, 50–59 (2018). doi: 10.1038/s41586-018-0176-1; pmid: 29875486
40. J. E. Huth, *The Lost Art of Finding Our Way* (Harvard Univ. Press, 2013).
41. D. Dreyer *et al.*, Bogong moths use a stellar compass for long-distance navigation at night. *Nature* **643**, 994–1000 (2025). doi: 10.1038/s41586-025-09135-3; pmid: 40533549
42. Materials and methods are available as supplementary materials.
43. P. A. LaChance, M. E. Hasselmo, Distinct codes for environment structure and symmetry in postrhinal and retrosplenial cortices. *Nat. Commun.* **15**, 8025 (2024). doi: 10.1038/s41467-024-52315-4; pmid: 39271679
44. A. S. Alexander *et al.*, Egocentric boundary vector tuning of the retrosplenial cortex. *Sci. Adv.* **6**, eaaz2322 (2020). doi: 10.1126/sciadv.aaz2322; pmid: 32128423
45. P.-Y. Jacob *et al.*, An independent, landmark-dominated head-direction signal in dysgranular retrosplenial cortex. *Nat. Neurosci.* **20**, 173–175 (2017). doi: 10.1038/nn.4465; pmid: 27991898
46. C. N. Boccara *et al.*, Grid cells in pre- and parasubiculum. *Nat. Neurosci.* **13**, 987–994 (2010). doi: 10.1038/nn.2602; pmid: 20657591
47. R. Chaudhuri, B. Gerçek, B. Pandey, A. Peyrache, I. Fiete, The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nat. Neurosci.* **22**, 1512–1520 (2019). doi: 10.1038/s41593-019-0460-x; pmid: 31406365
48. A. Rubin *et al.*, Revealing neural correlates of behavior without behavioral measurements. *Nat. Commun.* **10**, 4745 (2019). doi: 10.1038/s41467-019-12724-2; pmid: 31628322
49. Z. Ajabi, A. T. Keinath, X.-X. Wei, M. P. Brandon, Population dynamics of head-direction neurons during drift and reorientation. *Nature* **615**, 892–899 (2023). doi: 10.1038/s41586-023-05813-2; pmid: 36949190
50. J. D. Seelig, V. Jayaraman, Neural dynamics for landmark orientation and angular path integration. *Nature* **521**, 186–191 (2015). doi: 10.1038/nature14446; pmid: 25971509
51. A. Rubin, M. M. Yartsev, N. Ulanovsky, Encoding of head direction by hippocampal place cells in bats. *J. Neurosci.* **34**, 1067–1080 (2014). doi: 10.1523/JNEUROSCI.5393-12.2014; pmid: 24431464
52. S. Danilovich, Y. Yovel, Integrating vision and echolocation for navigation and perception in bats. *Sci. Adv.* **5**, eaaw6503 (2019). doi: 10.1126/sciadv.aaw6503; pmid: 31249874
53. G. G. Kwiecinski, T. A. Griffiths, Rousettus egyptiacus. *Mamm. Species* **199**, 1–9 (1999). doi: 10.2307/3504411
54. L. Las, N. Ulanovsky, Bats. *Nat. Methods* **21**, 1135–1137 (2024). doi: 10.1038/s41592-024-02330-6; pmid: 38997589
55. S. B. Childs, E. R. Buchler, Perception of simulated stars by *Eptesicus fuscus* (Vespertilionidae): A potential navigational mechanism. *Anim. Behav.* **29**, 1028–1035 (1981). doi: 10.1016/S0003-3472(81)80056-5
56. H. Mouritsen, O. N. Larsen, Migrating songbirds tested in computer-controlled Emlen funnels use stellar cues for a time-independent compass. *J. Exp. Biol.* **204**, 3855–3865 (2001). doi: 10.1242/jeb.204.22.3855; pmid: 11807103
57. M. Dacke, E. Baird, M. Byrne, C. H. Scholtz, E. J. Warrant, Dung beetles use the Milky Way for orientation. *Curr. Biol.* **23**, 298–300 (2013). doi: 10.1016/j.cub.2012.12.034; pmid: 23352694
58. F. Raudies, M. P. Brandon, G. W. Chapman, M. E. Hasselmo, Head direction is coded more strongly than movement direction in a population of entorhinal neurons. *Brain Res.* **1621**, 355–367 (2015). doi: 10.1016/j.brainres.2014.10.053; pmid: 25451111
59. Y. Yovel, M. Geva-Sagiv, N. Ulanovsky, Click-based echolocation in bats: Not so primitive after all. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **197**, 515–530 (2011). doi: 10.1007/s00359-011-0639-4; pmid: 21465138
60. A. S. Etienne, K. J. Jeffery, Path integration in mammals. *Hippocampus* **14**, 180–192 (2004). doi: 10.1002/hipo.10173; pmid: 15098724
61. A. Cheung, S. Zhang, C. Stricker, M. V. Srinivasan, Animal navigation: The difficulty of moving in a straight line. *Biol. Cybern.* **97**, 47–61 (2007). doi: 10.1007/s00422-007-0158-0; pmid: 17520273
62. M. Geva-Sagiv, L. Las, Y. Yovel, N. Ulanovsky, Spatial cognition in bats and rats: From sensory acquisition to multiscale maps and navigation. *Nat. Rev. Neurosci.* **16**, 94–108 (2015). doi: 10.1038/nrn3888; pmid: 25601780
63. M. R. Loewen, "Laboratory measurements of the sound generated by breaking waves," thesis, Massachusetts Institute of Technology and Woods Hole Oceanographic Institution (1991).
64. G. Koay, R. S. Heffner, H. E. Heffner, Hearing in a megachiropteran fruit bat (*Rousettus aegyptiacus*). *J. Comp. Psychol.* **112**, 371–382 (1998). doi: 10.1037/0735-7036.112.4.371; pmid: 9861710
65. J. Zeil, Visual navigation: Properties, acquisition and use of views. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **209**, 499–514 (2023). doi: 10.1007/s00359-022-01599-2; pmid: 36515743
66. K. Asumbisa, A. Peyrache, S. Trenholm, Flexible cue anchoring strategies enable stable head direction coding in both sighted and blind animals. *Nat. Commun.* **13**, 5483 (2022). doi: 10.1038/s41467-022-33204-0; pmid: 36123333
67. J. P. Goodridge, P. A. Dudchenko, K. A. Worboys, E. J. Golob, J. S. Taube, Cue control and head direction cells. *Behav. Neurosci.* **112**, 749–761 (1998). doi: 10.1037/0735-7044.112.4.749; pmid: 9733184
68. R. S. Heffner, G. Koay, H. E. Heffner, Sound localization in an Old-World fruit bat (*Rousettus aegyptiacus*): Acuity, use of binaural cues, and relationship to vision. *J. Comp. Psychol.* **113**, 297–306 (1999). doi: 10.1037/0735-7036.113.3.297; pmid: 10497795
69. A. Bicanski, N. Burgess, Environmental anchoring of head direction in a computational model of retrosplenial cortex. *J. Neurosci.* **36**, 11601–11618 (2016). doi: 10.1523/JNEUROSCI.0516-16.2016; pmid: 27852770
70. Y. Yan, N. Burgess, A. Bicanski, A model of head direction and landmark coding in complex environments. *PLOS Comput. Biol.* **17**, e1009434 (2021). doi: 10.1371/journal.pcbi.1009434; pmid: 34570749
71. J. J. Knierim, H. S. Kudrimoti, B. L. McNaughton, Place cells, head direction cells, and the learning of landmark stability. *J. Neurosci.* **15**, 1648–1659 (1995). doi: 10.1523/JNEUROSCI.15-03-01648.1995; pmid: 7891125
72. R. Knight *et al.*, Weighted cue integration in the rodent head direction system. *Philos. Trans. R. Soc. London Ser. B* **369**, 20120512 (2013). doi: 10.1098/rstb.2012.0512; pmid: 24366127
73. R. Eilam-Altstädter, L. Las, M. P. Witter, N. Ulanovsky, *Stereotaxic Brain Atlas of the Egyptian Fruit Bat* (Academic Press, 2021).
74. S. Palgi *et al.*, Data for: Head-direction cells as a neural compass in bats navigating outdoors on a remote oceanic island, Zenodo (2025); https://doi.org/10.5281/zenodo.15646845.

ACKNOWLEDGMENTS

We thank D. Derdikman, A. Finkelstein, Y. Ziv, A. Rubin, M. Kirschhock, R. Yerushalmi, and E. Farber for comments on the manuscript; S. Zikankuba, M. Mwanzalila, A. Massa, K. Kingojo, C. Haule, L. Rufino, M. Martinengo, J. Alexiou, G. Ginosar, M. Weinberg, K. Krivoruchko, M. Guggenberger, S. Forer, G. Omarch, S. Stephen, F. Essajee, H. Essajee, E. Kopf, S. Molcho, and S. Mazaki for assistance in the fieldwork, boat transportation to the island, work at TVLA, and logistical support; A. Shalev, N. Tzuri, M. Brezins, A. Pilewski, and G. Cohen for help in bat training and assistance in Israel; B. Pasmantirer, G. Ankaoua, and A. Jahanfard for mechanical designs and manufacturing; R. Tumasus, C. Ra'anan, and R. Eilam for histology; M. P. Witter for advice on histological delineation of dorsal presubiculum and retrosplenial cortex; and G. Brodsky for graphics. **Funding:** N.U. is the incumbent of the Barbara and Morris Levinson Professorial Chair in Brain Research. This study was supported in part by research grants to N.U. from the European Research Council (ERC-Advanced NEURO_OUTDOORS and ERC-Consolidator NATURAL_BAT_NAV); by research grants to N.U. and L.L. from the National Institutes of Health (NIH R01 – NS121413), CRCNS NSF-BSF (BSF 2020806), and the Israel Science Foundation (ISF 1829/23); by research grants to N.U. from the Weizmann Center for the Development of Innovative Neurotechnologies, Zuckerman Center for Learning Memory & Cognition, Center for Research on Perception and Action, Dr. Lou Siminovitch Laboratory for Research in Neurobiology, Dorraine S. Schwartz, Debra and Paul Sagues, and Yehuda and Judith Bronicki; by a research grant to H.M. from the European Research Council (ERC-SyG–Quantum Birds); a Deutsche Forschungsgemeinschaft collaborative grant to N.U. and H.M. (DFG no. 395940726, SFB 1372 "Magnetoreception and navigation in vertebrates"); and by the André Deloro Prize for Scientific Research and the Kimmel Award for Innovative Investigation to N.U. S.P. was supported by an Azrieli Brain Science fellowship for excellent PhD students; and S.R. was supported by postdoctoral fellowships from the European Molecular Biology Organization (ALTF 853-2017) and the Human Frontier Science Program (LT000365-2018L). **Author contributions:** N.U. conceived of this study in 2018 while scuba diving on a remote oceanic island. S.P., S.R., S.R.M., L.L., and N.U. designed the experiments. S.P., S.R., S.R.M., T.E., L.L., and N.U. developed technology, set up experimental systems, and conducted technical preparations for the island experiments. J.D.K. and A.I.A. obtained research permits and assisted with logistics. S.P., S.R., A.T., C.C., L.L., and N.U. performed preparations on the mainland and conducted the experiments on the island. S.P. performed the data analysis of the island experiments, and S.R., Y.W., and L.B.-A. contributed to the data analysis. S.R.M. and H.M. conducted the magnetic manipulation experiments in the tunnel, and S.R.M. analyzed the data from these experiments. N.U. guided all the data analysis, with major input from L.L. S.P and N.U. wrote the first draft of the manuscript, with major input from S.R., S.R.M., and L.L. All authors participated in writing and editing the manuscript. N.U. supervised the project. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The data and code that support the conclusions of this study are freely accessible online at Zenodo (*74*). **License information:** Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. This research was funded in whole or in part by the European Research Council (ERC-Advanced NEURO_OUTDOORS, ERC-Consolidator NATURAL_BAT_NAV, and ERC-SyG–Quantum Birds); as required the author will make the Author Accepted Manuscript (AAM) version available under a CC BY public copyright license.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adw6202
Materials and Methods; Supplementary Results; Figs. S1 to S20; References (*75*–*99*); MDAR Reproducibility Checklist

Submitted 10 February 2025; accepted 22 August 2025

10.1126/science.adw6202



AAAS.ORG/COMMUNITY







AAAS' Member Community is a one-stop destination for scientists and STEM enthusiasts alike. It's "Where Science Gets Social": a community where facts matter, ideas are big and there's always a reason to come hang out, share, discuss and explore.

Member COMMUNITY
AAAS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

RESEARCH ARTICLE SUMMARY

GLOBAL SEA LEVEL

Global mean sea level over the past 4.5 million years

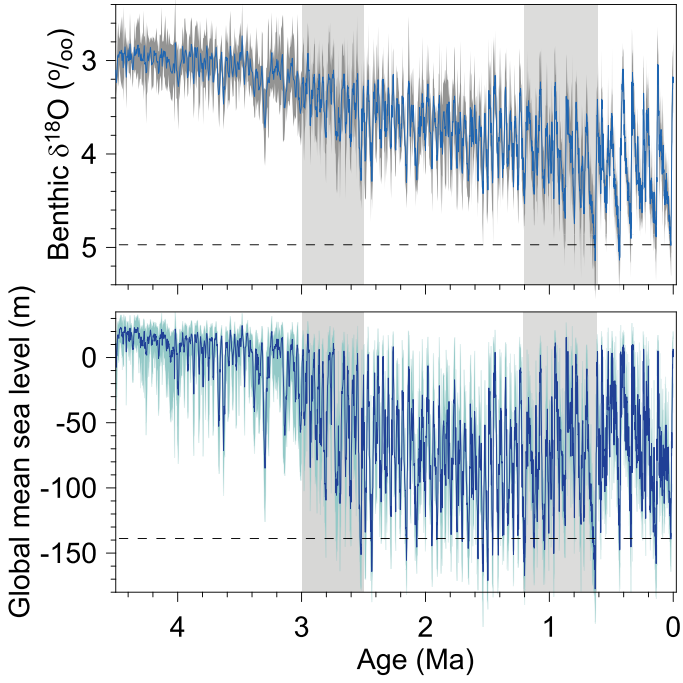
Peter U. Clark*, Jeremy D. Shakun, Yair Rosenthal, David Pollard, Steven W. Hostetler, Peter Köhler, Patrick J. Bartlein, Jonathan M. Gregory, Chenyu Zhu, Daniel P. Schrag, Zhengyu Liu, Nicklas G. Pisias

Full article and list of author affiliations: https://doi.org/10.1126/science.adv8389

INTRODUCTION: The oxygen isotopic composition ($^{18}O/^{16}O$) of benthic foraminifera shells (expressed as $\delta^{18}O_b$) has long provided a simple but powerful record of the long-term evolution of combined changes in deep ocean temperature (expressed as $\delta^{18}O_T$) and global mean sea level, with the latter recorded by changes in the $\delta^{18}O$ of seawater (expressed as $\delta^{18}O_{sw}$) due to ice sheet growth and decay. More recently, a widely used strategy to isolate $\delta^{18}O_{sw}$ by regressing the $\delta^{18}O_b$ record against times of known sea level reinforced the longstanding inference from the $\delta^{18}O_b$ record that sea level change during the past 4.5 million years (Myr) experienced two transitions towards lower sea level lowstands. The first transition [3 to 2.5 million years ago (Ma)] represents increasingly larger Northern Hemisphere (NH) ice sheets that fluctuated with a 41-thousand-year (kyr) periodicity. The second middle Pleistocene transition (MPT) (1.2 to 0.62 Ma) is characterized by even larger fluctuations of NH ice sheets, which changed from a 41-kyr periodicity to a dominant 100-kyr periodicity. In both cases, the transitions occurred in the absence of any changes in the orbital forcing by the Sun, suggesting that the cause(s) of these transitions was internal to the climate system.

RATIONALE: Despite general agreement that each of the two transitions represented an increase in the size of NH ice sheet fluctuations, the regression-based method to reconstruct sea level remains uncertain because by default, it reproduces the $\delta^{18}O_b$ variability. Indeed, multiple lines of evidence—including from terrestrial, marine, and geophysical records that constrain ice sheet extent, as well as from $\delta^{18}O_{sw}$ which is derived by subtracting known $\delta^{18}O_T$ from the $\delta^{18}O_b$ record—suggest that fluctuations of large NH ice sheets have occurred throughout the past 2.5 Myr, with the MPT thus recording a change in periodicity but not in ice sheet size. If correct, this scenario poses new challenges in understanding the 41-kyr periodicity of large ice sheets in a warmer world as well as the change in ice sheet variability during the MPT.

RESULTS: We converted a new $\delta^{18}O_{sw}$ record to sea level by applying a mass-balance approach that accounts for time-varying temperature and ice volume effects on the $\delta^{18}O$ of ice sheets. We find that the first transition in the $\delta^{18}O_b$ record represents an increase in NH ice sheet fluctuations to a size that then reoccurred throughout the remaining 2.5 Myr whereas the second transition represents a decrease in mean ocean temperature accompanied by an increase in its variability. Our climate model results show that the height of large pre-MPT ice sheets allowed them to be in surface mass balance under warmer-than-present global temperatures. Growth and decay of these large ice sheets was paced by 41-kyr obliquity forcing, with deglaciation occurring once ice sheets exceeded a certain size and became unstable. We propose that changes in the Southern Ocean carbon cycle during the MPT modulated the response of global temperature and ice sheets to 41-kyr obliquity forcing, resulting in the emergence of the ~100-kyr signal.



The benthic $\delta^{18}O$ record and global mean sea level over the past 4.5 Myr. The $\delta^{18}O$ record measured in benthic foraminifera over the past 4.5 Myr shows two transitions of increasing values (vertical gray rectangles) that are widely interpreted to represent increasingly large ice sheet fluctuations, whereas our new reconstruction of global mean sea level shows that this only applies to the first transition, with large fluctuations of ice sheets reoccurring over the past 2.5 Ma. Horizontal dashed lines are the benthic $\delta^{18}O$ value (upper panel) and the sea-level lowstand (lower panel) from the Last Glacial Maximum (21 ka).

CONCLUSION: Our work creates a new paradigm for the development of the Pleistocene ice ages, in particular the idea that large changes in sea level occurred throughout the Pleistocene rather than temperature varying with ice volume as the Earth cooled. In doing so, our findings modify as well as add several fundamental challenges to our understanding of ice sheet–climate interactions. Underlying each of these challenges is the fact that the dominant orbital-scale sea level variability and its changes over the past 3 Myr are not the ones that would be predicted solely by the associated orbital forcing, suggesting internal feedbacks of the climate system that we propose are largely driven by changes in the Southern Ocean carbon cycle and the effects of these changes on CO_2 and global temperature. High-resolution ice core CO_2 records that extend beyond 0.8 Ma are needed to test our hypotheses. □

*Corresponding author. Email: clarkp@onid.orst.edu Cite this article as P. U. Clark. *et al.*, *Science* **390**, eadv8389 (2025). DOI: 10.1126/science.adv8389

GLOBAL SEA LEVEL

Global mean sea level over the past 4.5 million years

Peter U. Clark[1,2]†*, Jeremy D. Shakun[3]†, Yair Rosenthal[4,5], David Pollard[6], Steven W. Hostetler[1], Peter Köhler[7], Patrick J. Bartlein[8], Jonathan M. Gregory[9,10], Chenyu Zhu[11], Daniel P. Schrag[12], Zhengyu Liu[13], Nicklas G. Pisias[1]

Changes in global mean sea level (GMSL) during the late Cenozoic remain uncertain. We use a reconstruction of changes in $\delta^{18}O$ of seawater to reconstruct GMSL since 4.5 million years ago (Ma) that accounts for temperature-driven changes in the $\delta^{18}O$ of global ice sheets. Between 4.5 and 3 Ma, sea level highstands remained up to 20 m above present whereas the first lowstands below present suggest onset of Northern Hemisphere glaciation at 4 Ma. Intensification of global glaciation occurred from 3 Ma to 2.5 Ma, culminating in lowstands similar to the Last Glacial Maximum lowstand at 21,000 years ago and that reoccurred throughout much of the Pleistocene. We attribute the middle Pleistocene transition in ice sheet variability (1.2 Ma to 0.62 Ma) to modulation of 41-thousand-year (kyr) obliquity forcing by an increase in ~100-kyr CO_2 variability.

The record of $\delta^{18}O$ measured in benthic foraminifera shells ($\delta^{18}O_b$) (Fig. 1A) is widely used to represent the long-term evolution of combined changes in ocean temperature and global mean sea level (GMSL) over the Cenozoic (*1*–*5*) as well as to evaluate the response of the climate system to forcings on orbital and tectonic timescales (10^4 to 10^6 years) (*6*–*9*). The fidelity of this record as a global signal has improved with the stacking of multiple records that increase the signal-to-noise ratio (*10*–*12*) and the use of several independent dating methods that provide a robust chronology (*1*, *11*, *13*, *14*). Different strategies have then been used to decompose the $\delta^{18}O_b$ record into its temperature ($\delta^{18}O_T$) and seawater ($\delta^{18}O_{sw}$) components, with the former reflecting the temperature that the benthic foraminifera shells formed in (larger $\delta^{18}O_T$ in lower temperature) and the latter largely reflecting global ice volume (larger $\delta^{18}O_{sw}$ when more land ice is present). Although these strategies share the same goal, their different underlying assumptions have resulted in solutions with pronounced differences during the late Pliocene, early Pleistocene [beginning 2.6 million years ago (Ma)], and middle Pleistocene transition (MPT) (defined herein as 1.2 to 0.62 Ma), thus confounding understanding of these critical periods of late-Cenozoic climate and GMSL change (fig. S1).

We address these issues by reconstructing the GMSL of the past 4.5 million years (Myr), which circumvents several shortcomings associated with previous reconstructions [Materials and Methods and supplemental materials (SM)]. Using our new reconstruction of changes in global mean ocean temperature (ΔMOT), we previously decomposed a $\delta^{18}O_b$ stack (*12*) that includes a correction for some possible combination of diagenesis (*15*, *16*) and the carbonate-ion effect on foraminifera (*17*) into its $\delta^{18}O_T$ and $\delta^{18}O_{sw}$ components (*18*). We convert our $\delta^{18}O_{sw}$ reconstruction (fig. S2C) to GMSL by applying a mass-balance approach that conserves $\delta^{18}O$ changes of the global ocean and the North American (Laurentide and Cordilleran) (NAIS), Eurasian (EUR), Greenland (GRN), and Antarctic (ANT) ice sheets, including loss of marine-based ANT ice (*19*, *20*) (see SM). In doing so, our model accounts for time-varying temperature (*21*) and ice-volume effects on the $\delta^{18}O$ of ice sheets ($\delta^{18}O_i$) over the past 4.5 Myr, yielding a variable relationship between $\delta^{18}O_{sw}$ and sea level (fig. S3), in contrast to methods that convert $\delta^{18}O_{sw}$ to sea level using a constant linear scaling between $\delta^{18}O_{sw}$ and sea level (*22*–*26*) or a variable scaling developed for only the last glacial cycle (*27*, *28*).

Global mean sea level over the past 4.5 Myr

Our new reconstruction (Fig. 1C) shows that GMSL was largely above present from 4.5 to 4.0 Ma, with sea level highstands of ~20 m and sea level fluctuations of 5 to 20 m, suggesting that contributions were likely from some combination of changes in GRN and ANT ice sheets although contributions from NAIS or EUR ice sheets cannot be excluded. Subsequent highstands between 4 and 3 Ma remained 15 to 20 m above present but lowstands began to extend regularly below present by as much as −50 to −80 m thus marking the onset of extensive Northern Hemisphere (NH) (NAIS, EUR) glaciation at ~4 Ma. The intensification of NH glaciation from 3 to 2.5 Ma was marked by highstands that decreased to near-modern levels and lowstands that progressively decreased from -55 ± 29 m at 3 Ma to -151 ± 25 m at 2.52 Ma, which is comparable to or lower than that of the Last Glacial Maximum (LGM) [21 thousand years ago (ka)] lowstand (-132 ± 15 m) (Fig. 1C). Since 2.5 Ma, sea level highstands have been within 10 m of present, except between 2.25 to 1.45 Ma when they remained as much as 50 m below present, suggesting incomplete deglaciation of NH ice sheets, while fluctuations of large LGM-size ice sheets continued throughout the Pleistocene.

Three factors used to derive our GMSL reconstruction combine to indicate that early Pleistocene ice sheets were comparable in volume to the LGM, in contrast to most previous reconstructions which suggest that such large ice sheets did not appear until after the MPT (fig. S6, A to D). First, with a ΔMOT that is ~50% of a corresponding change in global mean sea surface temperature (ΔSST) before the MPT (*18*), the increasing (cooling) trend of $\delta^{18}O_T$ is smaller in our reconstruction than if ΔMOT:ΔSST=1 is assumed throughout the Pleistocene. Being a residual from the observed $\delta^{18}O_b$, our $\delta^{18}O_{sw}$ is thus more negative during the Pliocene and has a larger increasing trend between 3.0 and 2.5 Ma (compare fig. S2A and S2B). Second, because ΔMOT:ΔSST was smaller during the early Pleistocene, so was glacial-interglacial MOT variability (*15*), limiting the contribution of $\delta^{18}O_T$ to glacial cycles in $\delta^{18}O_b$, in contrast to the middle and late Pleistocene when ocean temperature and ice volume exerted similarly sized controls on $\delta^{18}O_b$. Third, we find that higher global mean surface temperatures (GMST) during the early Pleistocene caused the ice sheets to be isotopically heavier (fig. S3D). This reduced their influence on $\delta^{18}O_{sw}$ relative to the isotopically lighter ice sheets during the middle and late Pleistocene. At the same time, given our reconstructed $\delta^{18}O_T$, (*18*) identified the need to remove a spurious increasing trend from the $\delta^{18}O_b$ stack attributed to diagenesis (*15*, *16*) or the carbonate-ion effect on benthic foraminifera (*17*) that if unaccounted for would have resulted in even more positive Pliocene $\delta^{18}O_{sw}$ values and larger early Pleistocene ice volumes than in our reconstruction (compare fig. S2B and S2C and fig. S2E and S2F).

The spectral characteristics of our GMSL reconstruction are dominated by 41-thousand-year (kyr) variance until ~1.2 Ma. This is followed by the full emergence of a ~100-kyr signal by 0.62 Ma (Fig. 2A)

[1]College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR, USA. [2]School of Geography and Environmental Sciences, University of Ulster, Coleraine, Northern Ireland. [3]Department of Earth and Environmental Sciences, Boston College, Chestnut Hill, MA, USA. [4]Department of Marine and Coastal Science, Rutgers State University, New Brunswick, NJ, USA. [5]Department of Earth and Planetary Sciences, Rutgers State University, New Brunswick, NJ, USA. [6]Earth and Environmental Systems Institute, Pennsylvania State University, University Park, PA, USA. [7]Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung, Bremerhaven, Germany. [8]Department of Geography, University of Oregon, Eugene, OR, USA. [9]National Center for Atmospheric Science, University of Reading, Reading, UK. [10]Met Office Hadley Centre, Exeter, UK. [11]Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China. [12]Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA. [13]Department of Geography, The Ohio State University, Columbus, OH, USA. ***Corresponding author. Email: clarkp@onid.orst.edu** †These authors contributed equally to this work.

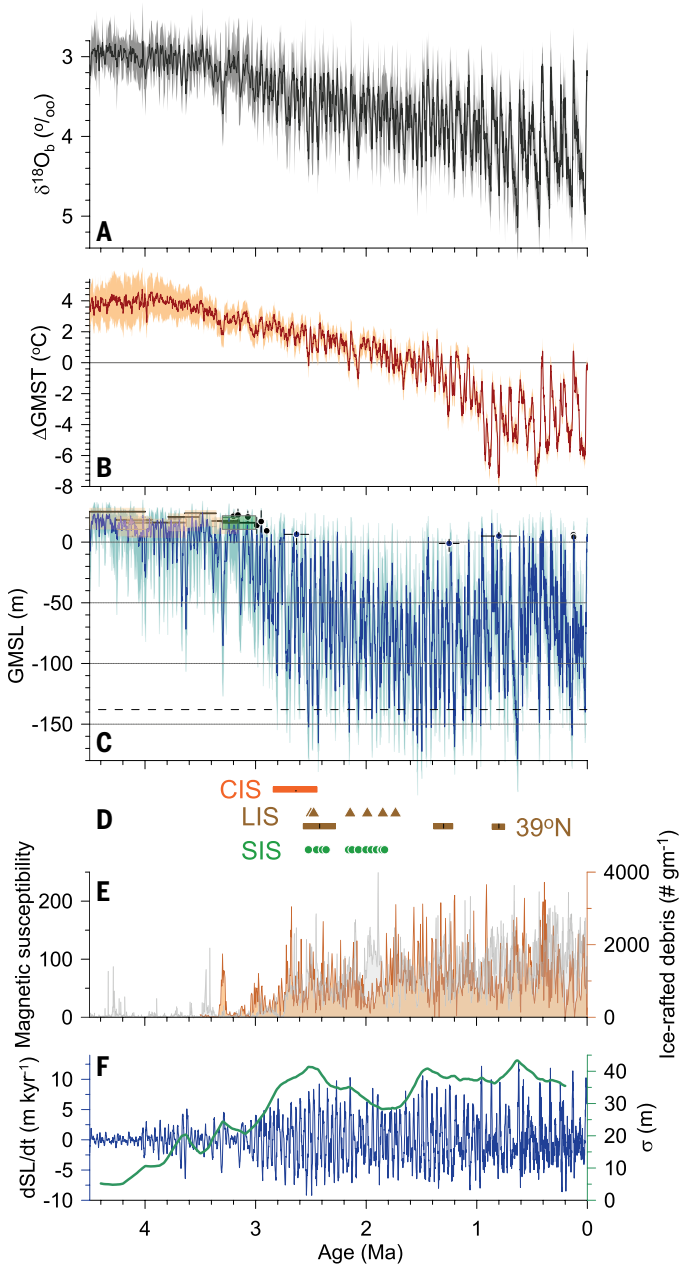


Fig. 1. Global mean temperature and sea level changes over the past 4.5 Myr. (**A**) Prob-stack $\delta^{18}O_b$ record (*12*). (**B**) Global mean surface temperature difference (ΔGMST) (dark red with 1σ uncertainty, shown in light orange) referenced pre-industrial (PI). (**C**) Global mean sea level (GMSL) (blue line with 1σ uncertainty, shown in light blue-green). The uncertainty on $\delta^{18}O_{sw}$ (converted to sea level) is calculated as the square root of the sum of squares of the uncertainty in the Prob-stack $\delta^{18}O_b$ record (*12*) and in the ΔMOT reconstruction (as $\delta^{18}O$) (*18*). Also shown are published Pliocene (*37*–*39*) and Pleistocene (*119*,*120*) sea level highstands. The dashed horizontal line is referenced to the LGM lowstand (−130 m). (**D**) Geologic evidence of extensive margins of early-to-middle Pleistocene Cordilleran Ice Sheet (CIS, orange symbols) (*43*), Laurentide Ice Sheet records at 39°N (LIS, brown symbols) (*42*, *44*), and Scandinavian Ice Sheet (SIS, green symbols) (*46*). Newer evidence suggests a more limited extent of the SIS in the North Sea prior to 1.1 Ma (*121*) than that inferred by (*46*), but the difference in extent and its contribution to GMSL is negligible. (**E**) Records of ice-rafted debris from the North Pacific (gray) (*48*) and Nordic Seas (orange) (*47*). (**F**) Rate of change (12-kyr running average) (dark blue line) and standard deviation (σ, 200-kyr centered) (green line) of GMSL.

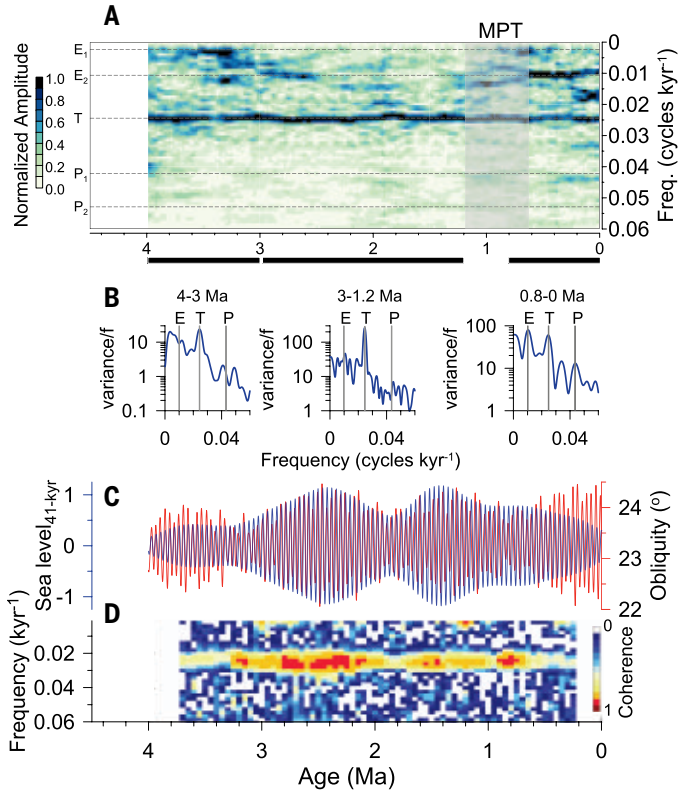


Fig. 2. Spectral characteristics of global mean sea level. (**A**) Evolutionary spectra of GMSL. The dominant periods in insolation variations at orbital periods are indicated by horizontal lines (eccentricity, $E_1 = 400$ kyr; $E_2 = 96$ kyr; obliquity, $T = 41$ kyr; precession, $P_1 = 23$ kyr, $P_2 = 19$ kyr). The vertical gray rectangle represents the MPT. (**B**) Spectral density of GMSL (variance per frequency) for three periods corresponding to the horizontal black bars shown in (A) (3 to 4 Ma, 1.2 to 3 Ma, 0 to 0.8 Ma). Orbital periods are indicated by vertical lines (E^2, eccentricity = 96 kyr; T, obliquity = 41 kyr; P^1, precession =23 kyr). Note scale change for *y*-axes (variance/f). (**C**) Filtered GMSL using a 41-kyr filter (blue line) and obliquity (red line). (**D**) Coherence between 41-kyr filtered GMSL and obliquity.

representing the ice sheet expression of the MPT. There have been similar concentrations of ~100-kyr and 41-kyr variance and a small concentration of 23-kyr variance over the past 0.8 Myr (Fig. 2, A and B), with the latter being consistent with a long-term increase in precessional forcing from eccentricity modulation (*29*) rather than increasing southward expansion of NH ice sheet margins during the MPT (*30*, *31*). A 41-kyr filtered sea level signal displays long-term (1 Myr) amplitude modulation of obliquity forcing until ~1 Ma, with the lowest coherence between both occurring during intervals of lower amplitude variations (i.e., nodes) (Fig. 2, C and D). The increase in amplitude after the first node centered on 3.3 to 3.1 Ma may reflect a threshold response to the increase in obliquity forcing that led to the intensification of NH glaciation (*32*). The below-present highstands and reduced variability between 2.25 Ma and 1.45 Ma (Fig. 1F) are associated with the second node at ~1.8 Ma (Fig. 2, C and D). The decrease in amplitude after the third node (~1 Ma) reflects the decrease in size of 41-kyr ice sheets relative to those that characterized the Pleistocene prior to 1 Ma, with large ice sheets now occurring less frequently (every ~100 kyr) (Fig. 1C).

The evolution of changes in sea level variability over the past 2.5 Myr as measured by GSML terminations (defined as the maximum rate of sea level rise during deglaciations) (Fig. 3D) shares the same relationship with obliquity as similarly defined GMST terminations (Fig. 3B). This includes the occurrence of terminations between 2.5 Ma

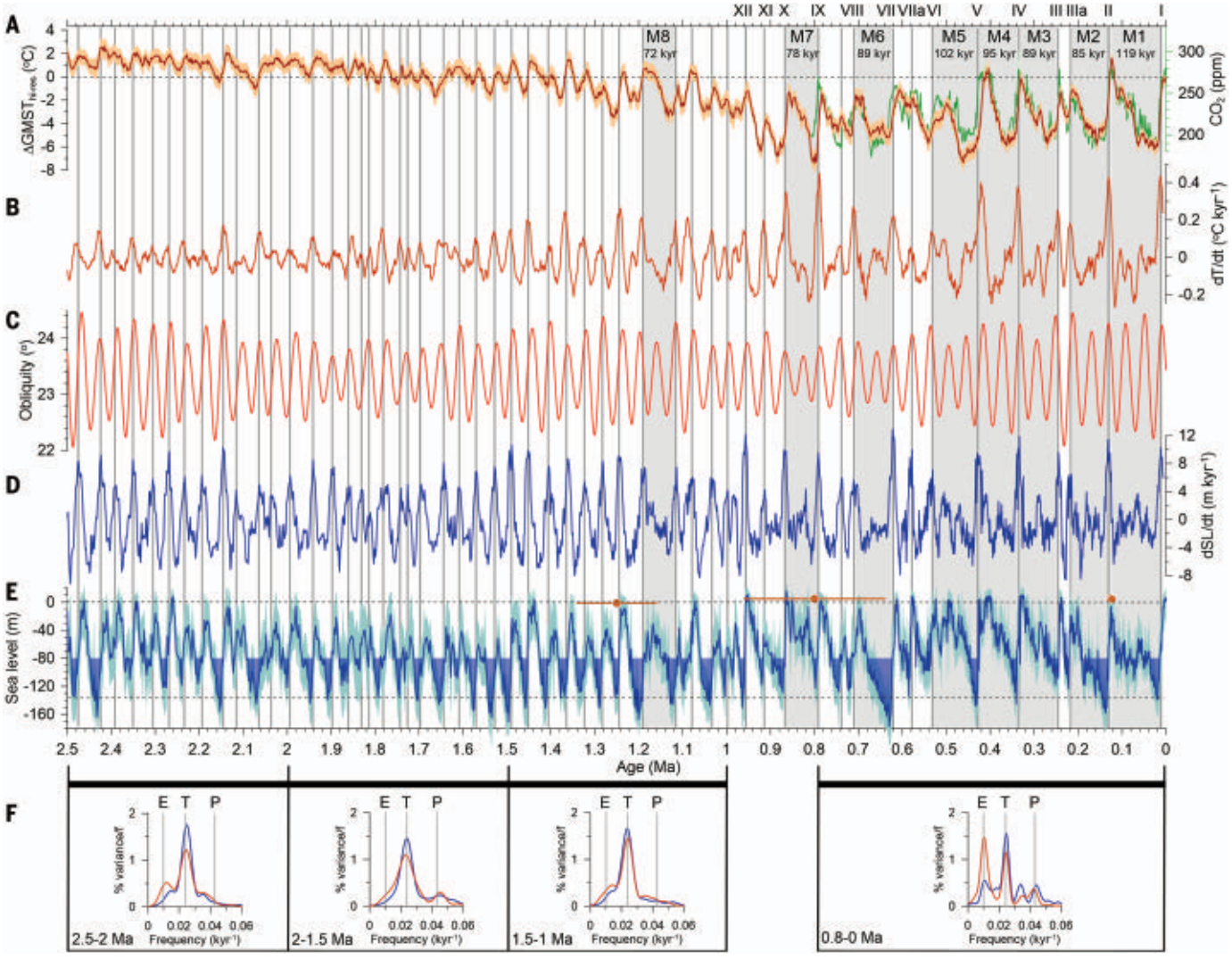


Fig. 3. Global temperature and sea level changes over the past 2.5 Myr. (**A**) Global mean surface temperature change (ΔGMST) (dark red with 1σ uncertainty, shown in light orange) referenced to PI, and atmospheric CO_2 (green) (*122*) placed on the AICC2023 chronology (*123*). (**B**) Rate of change of ΔGMST. (**C**) Obliquity (*71*). (**D**) Rate of change of GMSL (12-kyr running average), whose maxima are highlighted by vertical lines across (A) to (E). (**E**) GMSL; (blue line with uncertainty shown as light-blue green) sea level threshold of −80 m that when exceeded is followed by a termination upon next increase in obliquity (see text) is shown by filled blue sea level intervals. Dashed horizontal lines correspond to present (0 m) and LGM (−130 m) sea levels. Dark red symbols with 1σ age uncertainty are published Pleistocene sea level highstands (*119*,*120*). (**F**) Spectral density of rate of change of ΔGMST (dark red) and GMSL (blue) for four periods corresponding to horizontal black bars shown in (E). E, eccentricity = 100 kyr; T, tilt = 41 kyr; P, precession = 23 kyr. Frequency bands are noted by vertical lines.

and 1.2 Ma for every (near-) maximum in obliquity, as previously found with an independent age model (*13*), except during the node in obliquity modulation centered on 1.8 Ma (Fig. 2, C and D), when a small concentration of precessional variance appears (Fig. 3F).

Over the past 1.2 Myr, we identify eight intervals (M8 to M1) that begin and end with GMSL terminations associated with rising or peak obliquity (Fig. 3), with good agreement (within 0.5 to 4 kyr) between the mean ages of terminations during the past 1 Myr derived here and those derived from radiometric dating (*14*, *33*, *34*). Each of the eight intervals brackets one or two obliquity peaks without a corresponding termination (Fig. 3, D and E), consistent with previous work (*13*, *35*, *36*). This low-frequency signal first appears at 1.19 Ma during the 72-kyr M8 interval. This pattern is repeated for the 78-kyr M7 interval between Terminations X (0.87 Ma) and IX (0.79 Ma) and the 89-kyr M6 interval between Terminations VIII (0.71 Ma) and VII (0.62 Ma), with each of these two intervals followed by a shorter length interval when a termination accompanies an obliquity maximum. A similar interval of one skipped obliquity cycle between Terminations X and XII was inferred from a $\delta^{18}O_b$ record (*14*), but our ΔGMST and GMSL reconstructions show that the intervening Termination XI was as large as some earlier terminations (Fig. 3, B and D). Five more intervals with one or two skipped obliquity cycles (M5 to M1) span all of the past 0.53 Myr except for one brief 29-kyr interval between M3 and M2 (*34*), which reflects a temperature and sea level oscillation during Marine Isotope Stage 7 associated with the strongest obliquity forcing since 2.5 Ma (Fig. 3C). The average duration of the M7 to M1 intervals is 93.9 ± 13.4 kyr, resulting in the "100-kyr cycle" seen in the late-Pleistocene spectra of ΔGMST and GMSL (Fig. 3F) and their rates of change (Fig. 3, B and D). One notable difference is that GMSL terminations throughout the Pleistocene only occur after sea level falls below −80 m followed by the next increase to (near-) obliquity maxima (Fig. 3E), suggesting that ice sheets become unstable when

exceeding this size whereas GMST terminations have no comparable temperature threshold.

Our reconstruction of Pliocene highstands and the onset of NH glaciation between 4 and 3 Ma is in good agreement with other studies (*37*–*39*), including one that subtracted local temperature from $\delta^{18}O_b$ to derive $\delta^{18}O_{sw}$ and sea level as done here (*40*) (fig. S5). A similar local temperature-based reconstruction reproduces the intensification of NH glaciation that culminated in an LGM-like sea level lowstand at ~2.5 Ma (*26*) (fig. S5), while another reproduces large ice sheets throughout the Pleistocene (*41*) (fig. S6E). To first order, the similarity between temperature-based sea level reconstructions and our GMSL reconstruction is due to the local temperature reconstructions being nearly the same as our ΔMOT reconstruction (*18*). Several additional lines of evidence identify early Pleistocene NH ice sheets that were at or beyond their LGM extents by ~2.5 Ma, including well-dated terrestrial (*42*, *43*), marine (*44*, *45*), and geophysical (*46*) records (Fig. 1D) and ice-rafted debris records in the North Atlantic (*47*) and North Pacific (*48*) oceans (Fig. 1E), consistent with them being volumetrically as large as at the LGM.

By contrast, changes in Pleistocene GMSL inferred from the $\delta^{18}O_b$ record (*49*, *50*) (Fig. 1A) and from several reconstructions derived from it (*9*, *22*, *25*, *27*, *28*) find that the culminating lowstand at 2.5 Ma was only ~40% that of the LGM lowstand. Subsequent lowstands then remained at a similar intermediate level (~ −50 m) until the MPT when they furthered lowered, reaching LGM-like levels by 0.8 Ma (fig. S6, A to D). Given the robust evidence for spatially extensive early Pleistocene ice sheets (Fig. 1, D and E), the associated intermediate-size lowstands imply low-aspect ratio ice sheets relative to those following the MPT (i.e., thinner on average as the ice volume is less but the area is the same). This was the basis for the regolith hypothesis for the MPT (*51*). We attribute these findings of intermediate-size early Pleistocene ice sheets to the underlying methods that, by default, preserve the variability of the $\delta^{18}O_b$ record, including the increase in the size of glaciations during the MPT, with a further underestimation of early Pleistocene glaciations due to the lack of accounting for the long-term decrease in GMST on $\delta^{18}O_i$ (fig. S3) (see SM). Lastly, the good agreement between our reconstruction and all those that cover the past 0.8 Myr (fig. S7) suggest a robust understanding of GMSL since the MPT.

Toward resolving the paradox of large early Pleistocene ice sheets

A well-known paradox posed by early Pleistocene GMSL reconstructions is that their spectra are dominated by 41-kyr variability despite peak summer insolation being dominated by precession (*52*). Coupled climate and ice sheet models have reproduced NH ice sheet 41-kyr variability when accounting for albedo feedbacks, the integrated summer insolation forcing, and thin ice sheets with ablation zones that remained poleward of 60°N (*36*, *53*, *54*). However, such ice sheets are inconsistent with the geologic evidence for their southern margins in North America reaching 39°N during the early Pleistocene (Fig. 1D) (*42*, *55*), and they would not dominate the variability seen in our reconstruction. Another hypothesis proposes that higher early Pleistocene temperatures caused the response of the ANT ice sheet to precessional forcing to be out of phase with NH ice sheets, thus cancelling the 23-kyr signal in sea level records (*56*). However, hemispheric ΔSST reconstructions show a dominant in-phase 41-kyr signal with virtually no precessional power (*21*) and this hypothesis cannot account for the large obliquity-driven NH ice sheets in our reconstruction which had about twice the volume as modeled by (*56*).

We propose that obliquity forced variations in the Southern Ocean carbon cycle during the early Pleistocene caused dominant $[CO_2]_{atm}$ variability at this frequency (*57*). At the same time, the dominant 41-kyr signal of Southern Ocean SSTs and the temperature effects of related changes in the position of the sea ice edge would have been advected through the shallow meridional circulation and upwelled in the equatorial Pacific (*58*, *59*), where they combined with the SST variability from CO_2 forcing to induce the observed dominant 41-kyr SST signal at these low latitudes (*21*, *60*). Observations (*61*) and modeling (*62*) show that changes in equatorial Pacific SSTs can strongly influence NH ice sheet surface mass balance (SMB), with the ice sheet changes then potentially acting as a positive feedback on equatorial Pacific SSTs (*63*).

Our results pose another dimension to this paradox in showing that large ice sheets existed throughout the Pleistocene whereas glacial GMSTs underwent a ~6°C long-term cooling (Fig. 1, B and C), with polar amplification causing an even greater amount of cooling experienced by the high-latitude ice sheets (*21*). This relationship is further complicated by the strong effect that ice sheets have on their own local climate (*64*). We propose that the effects of higher early Pleistocene temperatures on ice sheet SMB were mitigated by an associated increase in high-latitude precipitation directly from warming (Clausius-Clapeyron relationship) as well as from latitudinal shifts in the westerlies as SST gradients changed (*65*). We can then infer that while cooling decreased accumulation over glaciated regions, the same cooling reduced surface ablation and thus caused SMB to remain positive. Similar transient relationships between other factors affecting SMB such as the influence of clouds on ice sheet surface energy balance (*66*, *67*) or terrestrial ecosystem emissions and atmospheric chemistry (*68*) can also be considered. If involved, changes in these transient feedbacks with global cooling continued to combine in such a way as to induce ice sheet SMB that supported the growth of large ice sheets and then maintained their presence in response to obliquity forcing, suggesting an extraordinary balancing act that allowed the inception and growth of ice sheets to their maximum size under a range of GMST and $[CO_2]_{atm}$.

To explore this issue further, we conducted several experiments with the GENESIS V3.0 climate model (*69*, *70*) to examine the sensitivity of early Pleistocene NH ice sheet SMB to orbits, atmospheric CO_2, and ice sheet height (see Materials and Methods and table S1). We first assessed our modeling strategy by running an experiment with LGM boundary conditions and found that the simulated NH ice sheets are in or have positive SMB. Our sensitivity experiments for early Pleistocene ice sheets included various combinations of orbits corresponding to low (at 2.165 Ma) and high (at 2.145 Ma) obliquity (*71*), $[CO_2]_{atm}$ of 200 and 300 parts per million by volume (ppmv) that covers much of the range suggested by proxies (*72*), and three ice sheet configurations that assess sensitivity to ice sheet height for a given LGM area (table S1). Our results suggest that early Pleistocene ice sheets that were larger than LGM ice sheets could have been in SMB with low obliquity and $[CO_2]_{atm}$ of 300 ppmv. SMB becomes more negative as ice sheet height is lowered to LGM values unless there is a compensatory decrease in $[CO_2]_{atm}$. No decrease in $[CO_2]_{atm}$ within the range of our current understanding of its Pleistocene variability could support the low-aspect-ratio ice sheets in SMB inferred by the regolith hypothesis (*51*).

We thus conclude that the extra height of the early Pleistocene ice sheets relative to the LGM was critical to them being in SMB during glaciations. Given our finding that the large ice sheets shared a common size threshold throughout the Pleistocene equivalent to sea level below −80 m that, when exceeded, triggered ice sheet instabilities that culminated in a termination (Fig. 3E), we have narrowed this aspect of the paradox to understanding how different boundary conditions and feedbacks may have contributed to the inception and growth of the early Pleistocene ice sheets.

Toward resolving the paradox of the middle Pleistocene transition and the origin of the ~100-kyr cycle

The longstanding MPT paradox relates to the increase in amplitude and decrease in frequency of GMSL variability during the MPT in the

absence of any corresponding change in orbital forcing (*4*). Our finding that large ice sheets existed throughout the Pleistocene, however, challenges several key aspects of their behavior that are usually invoked to explain this paradox: (i) that large ice sheets developed during the MPT in response to long-term changes internal to the climate system (*4*, *51*, *73*–*78*); (ii) that the time constant for growth of large ice sheets is considerably longer than for ice sheet decay, giving rise to ~100-kyr cycles (*79*–*81*); and (iii) that through their large inertia and influence on climate, the large ice sheets drove the ~100-kyr cycle in the rest of the climate system (*8*, *82*, *83*).

We first note that because pre-MPT ice sheets were as large as those after the MPT, their albedo forcing should have remained largely the same throughout the Pleistocene and thus cannot explain the large increase in orbital-scale GMST variability during the MPT (Fig. 1B). Instead, we attribute that increase to an increase in CO_2 variability associated with changes in the Southern Ocean carbon cycle. (*21*) proposed that the gradual cooling of the Southern Ocean through the early Pleistocene initiated the MPT in GMST at ~1.5 Ma when average extratropical Southern Hemisphere SSTs first decreased below preindustrial levels, increasing the sensitivity of sea ice to the dominant obliquity forcing at these high latitudes. Increased cooling of the Southern Hemisphere also led to an equatorward shift of the westerly winds (*84*), an increase in dust flux to the Southern Ocean (*85*), and greater stratification of surface waters (*86*). Through their impacts on the Southern Ocean carbon cycle (*84*, *85*, *87*–*90*), these changes increased carbon storage during glaciations, thus increasing obliquity-scale variability of $[CO_2]_{atm}$ and GMST. However, a further decrease and then stabilization of Southern Ocean ΔSSTs at the end of the MPT in GMST ~0.8 Ma then muted the response of the Southern Ocean carbon cycle to obliquity forcing that may have dominated $[CO_2]_{atm}$ variability during the early Pleistocene, with an attendant large decrease in the concentration of 41-kyr variance in $[CO_2]_{atm}$ and GMST (*21*).

We similarly attribute the decrease in large–ice sheet 41-kyr variance across the MPT (Fig. 2, A and B) to modulation of obliquity forcing by the newly established decrease in $[CO_2]_{atm}$ and GMST variance. Specifically, prior to reaching their maximum size at the end of each of the M intervals, ice sheet response to low obliquity was modulated by periods of higher $[CO_2]_{atm}$ and GMST and vice versa, with times of maximum ice sheet growth only occurring when the next obliquity low combined with low $[CO_2]_{atm}$ and GMST (Fig. 4, A to C). The initial retreat of large ice sheets in response to the next increase in obliquity then triggered a sequence of events and feedbacks leading to a termination every ~100 kyr (*34*). Among these feedbacks are large–ice sheet instabilities that sustained ice sheet retreat (*8*, *80*, *91*–*94*) which, through its influence on the Atlantic meridional overturning circulation, induced warming of the Southern Ocean with an associated retreat of sea ice, deterioration of the Southern Ocean halocline, and poleward shift in the westerlies. These changes then led to the release of CO_2 from the deep ocean (*90*) which combined with obliquity (Fig. 4D) to fully deglaciate NH ice sheets and usher in the next interglaciation.

Contrasting temperature and ice sheet responses to CO_2 forcing over the past 0.8 Myr are clearly expressed in their variance spectra, with a dominant ~100-kyr cycle in GMST (*21*) that is shared by CO_2 (fig. S9A), whereas the concentration of GMSL variance is nearly equally distributed between 41 kyr and ~100 kyr (Fig. 2B and fig. S9B). Insofar as the variance characteristics of GMST are shared by GMSST, MOT, and thus $\delta^{18}O_T$ (*18*) and those of GMSL are shared by $\delta^{18}O_{sw}$, we can similarly evaluate the relative contributions of $\delta^{18}O_T$ and $\delta^{18}O_{sw}$ to the variance of the $\delta^{18}O_b$ record since the MPT. The $\delta^{18}O_b$ record has long been used to argue that the dominant post-MPT ~100-kyr signal in that record (Fig. 5A) is from ice sheet dynamics (*8*, *14*, *95*–*97*), but our results show that its origin is largely in the $\delta^{18}O_T$ component due to changes in the carbon cycle (Fig. 5B), as first recognized by (*78*). Overall, however, terminations present in both $\delta^{18}O_T$ and $\delta^{18}O_{sw}$ were set by the threshold of when ice sheets exceeded a limiting size beyond which they responded nonlinearly to obliquity forcing and helped drive a termination. This threshold was reached for nearly every obliquity cycle of the Pleistocene prior to the MPT (Fig. 3, D and E). Afterward, the development of the ~100-kyr cycle arising from the long-term cooling enhanced the ability of the Southern Ocean carbon cycle to sequester CO_2 and modulate the response of $\delta^{18}O_T$ and $\delta^{18}O_{sw}$ to obliquity forcing until the growth of the next large ice sheet.

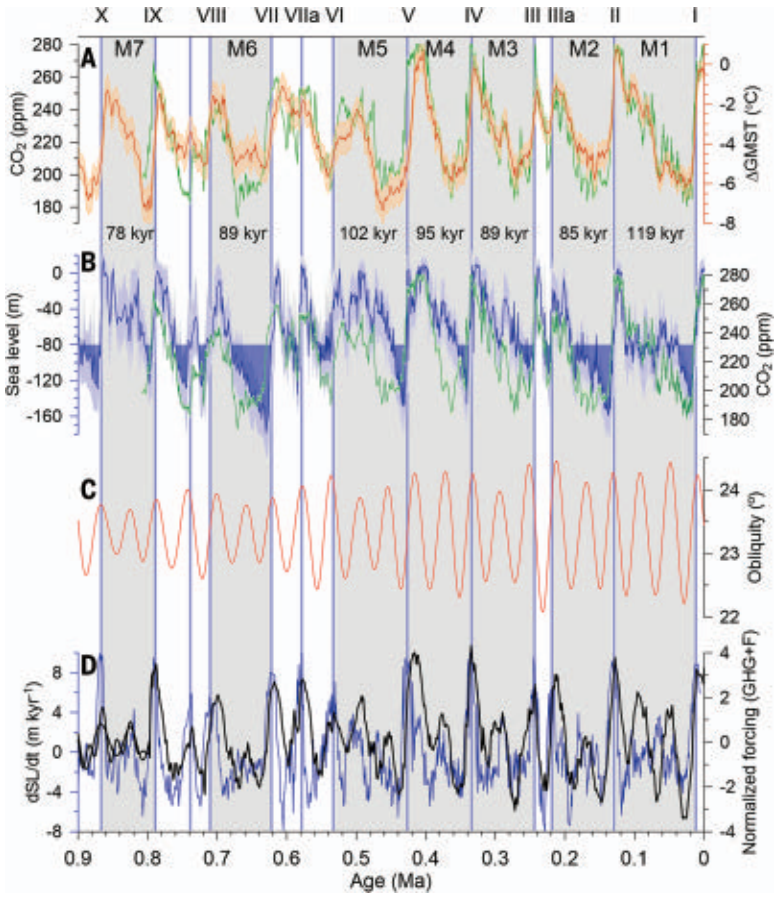


Fig. 4. Assessing the origin of the ~100-kyr cycle. (**A**) Global mean surface temperature change (ΔGMST) (dark red, with uncertainty shown in light orange) referenced to PI and atmospheric CO_2 (green) (*122*), both with 1σ uncertainty. (**B**) Global mean sea level (GMSL) (blue) and atmospheric CO_2 (green) (*122*), both with 1σ uncertainty (light blue and light green, respectively). (**C**) Obliquity (*71*). (**D**) Rate of change of GMSL (blue) compared with a normalized forcing function F (GHG+*F*) (*124*) that combines normalized atmospheric CO_2 from the Antarctic ice core record over the past 0.8 Myr (black) (*122*) placed on the AICC2023 chronology (*123*) and normalized obliquity. Note that there is no significant improvement in the correlation between rate of change of GMSL and a forcing function that combines various proportions of precession and obliquity (5:95, 25:75, 50:50).

Conclusions

We find that between 4.5 and 3 Ma, sea level highstands remained ~20 m above present, suggesting large decreases in Antarctic ice sheet

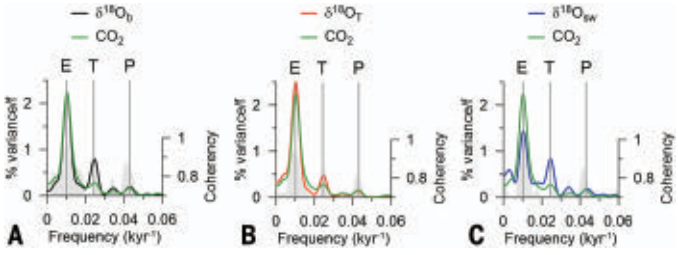


Fig. 5. Spectral density of the $\delta^{18}O_b$ record, its components, and atmospheric CO_2 for past 0.8 Myr. (**A**) Spectral density (normalized variance) of the Prob-stack $\delta^{18}O_b$ record (*12*) and atmospheric CO_2 (green) (*122*). (**B**) Spectral density (normalized variance) of the temperature component ($\delta^{18}O_T$) of the $\delta^{18}O_b$ record (*18*) and atmospheric CO_2 (green) (*122*). (**C**) Spectral density (normalized variance) of the seawater component ($\delta^{18}O_{sw}$) of the $\delta^{18}O_b$ record (*18*) and atmospheric CO_2 (green) (*122*).

and Greenland ice sheet volume relative to today. Beginning at 4 Ma, sea level lowstands began to episodically extend below present by as much as −50 to −80 m, suggesting the onset of NH glaciation. Sustained intensification of NH glaciation occurred from 3 Ma to 2.5 Ma, with highstands decreasing to near-modern levels while lowstands decreased to levels comparable to or lower than the LGM lowstand (~21 ka), which continued throughout much of the Pleistocene.

Our finding that large ice sheets existed throughout the Pleistocene modifies as well as adds several fundamental challenges to our understanding of ice sheet–climate interactions and their joint responses to external forcing. Underlying each of these challenges is the fact that the dominant orbital-scale GMSL variability and its changes over the past 3 Myr is not the one that would be predicted solely by the associated orbital forcing, suggesting internal feedbacks of the climate system that we propose are largely driven by changes in the Southern Ocean carbon cycle and their effect on $[CO_2]_{atm}$ and GMST. High-resolution ice core CO_2 records that extend beyond 0.8 Ma are needed to test our hypotheses. Finally, the presence of large ice sheets throughout the MPT indicates that the hypothesis of regolith removal to explain the change in ice sheet sensitivity to orbital forcing (*5*, *51*) can be rejected.

Materials and Methods

Assessment of GMSL reconstructions

Two different approaches have been used to decompose the $\delta^{18}O$ record measured in benthic foraminifera ($\delta^{18}O_b$) into its two primary components: the temperature ($\delta^{18}O_T$) and the $\delta^{18}O$ of seawater ($\delta^{18}O_{sw}$) that the benthic foraminifera shells formed in, with the latter largely reflecting GMSL. (*24*) developed an approach that regresses independently known sea level data (e.g., from corals) on $\delta^{18}O_b$ for the last glacial cycle and then uses this regression to reconstruct sea level from $\delta^{18}O_b$ records for the last four glacial cycles. Sea level is then converted to $\delta^{18}O_{sw}$ using a constant relation of $\Delta\delta^{18}O_{sw}$:ΔGMSL. (*98*) and (*25*) extended the regression over the last two glacial cycles to reconstruct sea level from $\delta^{18}O_b$ records over the last 5 Myr (figs. S1B and S6C). (*27*, *28*) further extended the regression over the last 800 kyr using the LR04 $\delta^{18}O_b$ stack and a stack of sea level records from (*99*) (fig. S7C) to reconstruct sea level over the last 40 Myr (figs. S1B and S6A show last 4.5 Myr). (*27*, *28*) also accounted for $\delta^{18}O$ variations in land ice ($\delta^{18}O_i$) over the last glacial cycle which they then applied to their sea level record to derive $\delta^{18}O_{sw}$ from $\delta^{18}O_b$, but this did not include the effect of increasing temperatures on $\delta^{18}O_i$ prior to 0.8 Ma. In any event, applying this regression approach to a $\delta^{18}O_b$ record will, by default, reproduce the variability of the $\delta^{18}O_b$ record, including the increase in the size of glaciations during the mid-Pleistocene transition (MPT), with the early Pleistocene low-amplitude variability further reinforced by not accounting for higher temperatures on $\delta^{18}O_i$.

(*9*) and (*100*) reconstructed sea level using a simple scaling relationship that determines the fraction of the LGM sea level lowstand (−120 m in their case) from the fraction of the $\delta^{18}O_b$ value at any given time relative to the LGM $\delta^{18}O_b$ value. As with the regression method, this relationship reproduces the $\delta^{18}O_b$ variability throughout the Pleistocene and does not account for the effect of increasing temperatures on $\delta^{18}O_i$ prior to 0.8 Ma, resulting in smaller early Pleistocene ice sheets than in our reconstruction (figs. S1B and S6D). Finally, these approaches imply a stationary partitioning of $\delta^{18}O_b$ between temperature and ice volume and thus do not capture the larger ice volume component relative to temperature that we identify before the MPT (*18*).

Another approach similar to ours subtracts independently reconstructed seawater temperature as $\delta^{18}O_T$ from $\delta^{18}O_b$ to derive $\delta^{18}O_{sw}$ (*22*, *23*, *41*, *101*–*103*), which can then be converted to GMSL using a relation of $\Delta\delta^{18}O_{sw}$:ΔGMSL (*22*, *23*). (*22*) used a reconstruction of bottom water temperature (BWT) similar to ours in showing lower ΔBWT than ΔSST during the early Pleistocene (*18*). However, this is a 2-Myr smoothed record and thus does not capture the increase in orbital variability across the MPT (*18*) (figs. S1B and S6B), which contributes to larger early Pleistocene ice sheets in our reconstruction. Ref. (*22*) also applied a constant 0.013‰ m^{-1} relationship to derive sea level from their $\delta^{18}O_{sw}$ which does not account for the effect of warmer early Pleistocene temperatures on $\delta^{18}O_i$, and thus $\delta^{18}O_{sw}$. Moreover, this relationship is more appropriate for warmer climates with just the Greenland and Antarctic ice sheets (*19*) as opposed to the 0.008‰ m^{-1} relationship that is more suitable for deriving sea level after 0.8 Ma. Accordingly, their sea level lowstands are generally higher than in our reconstruction throughout the Pleistocene (figs. S6B and S7B).

Other studies that used independent orbital-scale temperature records to derive $\delta^{18}O_{sw}$ from $\delta^{18}O_b$ suggest small GMSL variability in the late Pliocene that increases into the early Pleistocene and continues across the MPT (*26*, *40*, *41*). We used our method to convert $\delta^{18}O_{sw}$ data from (*41*) and (*26*) to sea level and find generally good agreement with our GMSL reconstruction (figs. S5 and S6E). However, the temperature records in (*41*) and (*26*) are local and so may not be representative of mean ocean temperature (MOT). Finally, other than (*27*) and (*28*), none of the reconstructions using either approach account for changing $\delta^{18}O_i$ on $\delta^{18}O_{sw}$ (*28*, *104*).

Finally, (*105*) and (*106*) reconstructed changes in water depth using grain size variations in uplifted marine sedimentary sequences preserved in New Zealand (figs. S1 and S8), with inferred water depths being supplemented by estimates from benthic foraminifera. Unconformities in the sedimentary sequence are assigned a maximum water depth of −32.5 m, and the authors assume the missing part of the sea level record to be <5 m. Water depth is converted to relative sea level using back stripping methods that account for sediment compaction and tectonic subsidence. Their reconstruction cannot be referenced to Holocene sea level, but it does identify long-term trends as well as the frequency and amplitude of sea level change on orbital timescales. Comparison of their reconstruction to other sea level reconstructions (fig. S1) and to our reconstruction (fig. S8) shows a similar concentration of 41-kyr variance but disagrees with these other reconstructions in showing no trend of increasing lowstands (i.e., greater amplitude) during the intensification of Northern Hemisphere glaciation.

Conversion of $\delta^{18}O_{sw}$ to sea level

We develop a simple model to translate $\delta^{18}O_{sw}$ to sea level that conserves global water and ^{18}O mass as they are redistributed between the ocean and the North American (NAIS) (Laurentide and Cordilleran ice sheets), Eurasian (EUR), Greenland (GRN), and Antarctic (ANT) ice sheets. The model also accounts for the changing isotopic composition of each ice sheet through time as a function of changes in its volume and global temperature, which allows us to separate their isotopic

effects. This is important because our reconstructions suggest that there is not a constant scaling between global temperature and ice volume over the past 4.5 Myr.

The method is based on conservation of global water and isotopic ^{18}O mass, extending similar treatments in (*19*) and (*20*). A simple expression is used for the mean $\delta^{18}O$ of each ice sheet as a linearized function of its size and global mean surface temperature (GMST), and a parameterization is assumed for the relative size of each ice sheet to the others at all times. After some tedious but straightforward algebra, this yields a quadratic equation yielding total ice volume over flotation (and hence global sea level rise) as a function of $\delta^{18}O_{sw}$ and GMST.

While $\delta^{18}O_{sw}$ is often converted to sea level using a constant scaling of ~-0.008 ‰/m based on LGM values [i.e., a ~1 ‰ $\delta^{18}O_{sw}$ enrichment (*107*) and ~130 m sea level lowering (*108*)], our model accounts for time-varying temperature and ice-volume effects on ice sheet $\delta^{18}O_i$ (fig. S3D), yielding a variable relationship between $\delta^{18}O_{sw}$ and sea level (fig. S3, A and B). As a result, relative to linear scaling, applying our model to our $\delta^{18}O_{sw}$ reconstruction produces lower sea levels during early Pleistocene glaciations because less-negative $\delta^{18}O_{ice}$ then (due to higher temperatures) requires greater ice growth to explain $\delta^{18}O_{sw}$ enrichments, and thus lower sea levels during Pliocene interglaciations because more-negative $\delta^{18}O_{ice}$ then (due to NAIS and EIS absence) requires less ice loss to explain $\delta^{18}O_{sw}$ depletions (fig. S3A).

In the SM, we explain the details of our model. Section 1 derives separate equations for differing climates relative to today. Section 2 outlines how we estimate the coefficients in the expression for mean ice sheet $\delta^{18}O$ based on observed quantities and published general circulation model (GCM) experiments. Section 3 explains the scheme we use to account for the residence time of ice in each ice sheet, and a list of prescribed quantities is given in section 4. Finally, section 5 assesses the sensitivity of our sea level reconstruction to the various parameters in our model. Finally, section 6 provides the MATLAB code for converting $\delta^{18}O_{sw}$ to sea level.

Time series analyses

We characterized the temporal variability and covariability of the different time series in four ways, using: 1) evolutionary spectra (Fig. 2A); 2) "global" multitaper spectral analysis (Fig. 2B); 3) bandpass-filtered time series of ΔGMST and GMSL at the 41-kyr obliquity period (Fig. 2C); and 4) the rolling coherence between the filtered series and obliquity (Fig. 2D). Evolutionary spectra (*109*) were calculated using the R astrochron package (*110*, *111*). The function eha() in the astrochron package implements a moving-window application of the multitaper method (MTM) spectral-analysis procedure (*112*). We used a window width of 500 kyr, a time step of 10 kyr between adjacent window positions, and a "standard" MTM time-bandwidth product of 4. Prior to analysis, the time series were detrended using a smoothing spline with 10 degrees of freedom using the spline.smooth() function in the R base stats package (*111*). This detrending step removes the very long-period variability from the time series but does not obscure or overly emphasize the eccentricity band variations in the data. The spectral density was rescaled to "normalized amplitude" values—in each window, the amplitudes (the square root of the spectral density) at each frequency were divided by the maximum amplitude across all frequencies. This facilitates judging the relative importance of variations across periods. We did not do any significance testing of the results because the presence of variability in the time series at orbital frequencies is really not in question, and white-noise or low-order autoregressive null-hypothesis spectra are rather naïve. On Fig. 2A, the dominant periods in insolation variations at orbital periods are indicated by horizontal lines (eccentricity: $E_1 = 400$ kyr, $E_2 = 96$ kyr; obliquity: $T = 41$ kyr and precession: $P_1 = 23$ kyr, $P_2 = 19$ kyr).

Global (as opposed to evolutionary) spectral analysis was performed for blocks of data, as indicated by the black horizontal bars on Fig. 2A. We used the multitaper method here as well. We bandpass-filtered the time series at the 41-kyr obliquity period using the butter() (Butterworth filter) function from the R signal package (*113*). The band limits were 39.0 and 43.1 kyr, 0.95 and 1.05 times the obliquity frequency of 1/41.0 = 0.0244 cycles kyr^{-1}. The coherence between obliquity and ΔGMST and GMSL was calculated using a moving window approach. Cross-spectral analyses were done following the Blackman-Tukey method using the SPECTX2 function from the ARAND software package (*114*). The number of lags used correspond to ~1/3 the length of the interval analyzed.

Ice sheet surface mass balance

We assessed the sensitivity of the surface mass balance (SMB) of early Pleistocene Northern Hemisphere ice sheets to orbits, atmospheric CO_2, and ice sheet height by conducting experiments with the GENESIS V3.0 climate model (*69*, *70*) (table S1). The GENESIS climate model comprises a T31 atmospheric model (horizontal resolution of approximately 3.75° × 3.75° latitude by longitude) coupled with the land-surface model LSX (2° × 2° resolution). SSTs were computed by the 50-m mixed layer ocean model in GENESIS. SMB is computed from annual precipitation, runoff, and evaporation over the ice sheets on the 2° × 2° grid of the LSX land surface model. The fraction of runoff retained through refreezing is computed as a function of temperature and precipitation using the method of (*115*). We ran the simulations for 100 years and analyzed the last 70 years to exclude spin-up artifacts.

Our sensitivity experiments used orbits corresponding to low (glacial) and high (interglacial) obliquity at 2.165 Ma and 2.145 Ma (*71*), respectively, and atmospheric CO_2 values of 200 and 300 ppmv which cover much of the early Pleistocene range suggested by proxies (*57*, *116*, *117*). We used three ice sheet configurations to assess sensitivity to ice sheet height for a given LGM area (table S1). The GLAC-1D LGM ice sheets are from (*118*). The HIGH ice sheets are intended to represent the largest of the early Pleistocene ice sheets, with ~30 m of additional lowering of sea level relative to the LGM sea level lowstand (Figs. 1C and 3E). We derived the HIGH ice sheet heights by calculating the volume of 30 m of global sea level and distributing the height equivalent of the volume uniformly over the GLAC 1-D LGM North American (NA) (20 m) and Eurasian (EU) (10 m) ice sheet complexes, accounting for the density ratio of water to ice (1028 kg m^{-3}/910 kg m^{-3}). The LOW ice sheet was derived by reducing the GLAC 1-D LGM ice sheet height to an equivalent of 60% of the full LGM sea level volume and is intended to represent the low-aspect ratio ice sheet geometry inferred from the regolith hypothesis (*51*).

Experiment EXP1 shows the model's ability to simulate NH (GLAC-1D) ice sheets that are in SMB or have positive SMB under LGM boundary conditions (table S1). Experiments EXP2 and EXP3 assess the effect of orbital changes on NH ice sheet SMB (table S1). In both experiments, NH surface temperatures north of 40°N are 8 K to 9 K colder than the PI due to the presence of the ice sheets while North Atlantic SSTs are 3 K to >5 K colder due to advection of cold air off the NA ice sheet complex and southward expansion of sea ice (fig. S10A). Warming of the oceans elsewhere partially offsets the NH cooling such that global SSTs are the same as PI and GMSTs are 2 K colder than PI (table S1). The net effect of the boundary conditions yields a glacial NA ice sheet-complex SMB of 5 m yr^{-1} (fig. S10B) and an interglacial SMB of −11 m yr^{-1}. Comparison of the NA and EU ice sheet-complex SMB in EXP2 to that of EXP3 indicates that glacial-interglacial obliquity accounts for 16 m yr^{-1} over NA and 2 m yr^{-1} over EU (table S1 and fig. S10C).

Experiments EXP2 and EXP4 assess the effect of CO_2 changes on NH ice sheet SMB. Over the 200 to 300 ppmv range evaluated here, CO_2 accounts for 15 m yr^{-1} in the NA ice sheet complex SMB and 8 m yr^{-1} in the EU ice sheet complex SMB, which are comparable to the effect of obliquity (table S1). Experiments EXP2 and EXP5 and experiments EXP4 and EXP6 assess the effect of ice sheet height on NH ice sheet SMB (table S1). Lowering the ice sheet height from HIGH (EXP2) to GLAC-1D (EXP5) results in SMB becoming more negative by 35 m yr^{-1} for the NA ice sheet complex and 11 m yr^{-1} for the EU ice sheet complex

(table S1). The effect of height on mass balance is most evident along the margin of the NA ice sheet complex (fig. S10D) where the higher margin blocks penetration of precipitation into the interior (orographic shadowing) and positions more ice area at or above the zero-degree isotherm. The marginally more positive SMB over the interior of the NA ice sheet complex for the lower ice sheet where precipitation is slightly greater than that of the high ice sheet reflects the effect of desertification.

Lowering the ice sheet height from HIGH (EXP4) to LOW (EXP6) results in SMB becoming more negative by 84 m yr^{-1} for the NA ice sheet complex and by 4 m yr^{-1} for the EU ice sheet complex (table S1). We note that the strongly negative SMB of the LOW ice sheet in EXP6 (table S1) indicates that such an ice sheet could not have existed even under glacial orbitals and LGM CO_2 levels.

In summary, our sensitivity tests suggest that early Pleistocene ice sheets that were larger than LGM ice sheets could have been in SMB with low obliquity and CO_2 levels of 300 ppmv. SMB becomes more negative as ice sheet height decreases toward LGM values unless there is a compensatory decrease in CO_2, but no decrease in CO_2 within the range of our current understanding of its Pleistocene variability could support the low-aspect-ratio ice sheets in SMB inferred by the regolith hypothesis (*51*). We note that while our experiments using LSX and a mixed layer ocean model are a first-order assessment of the sensitivity of ice sheet SMB to changes in boundary conditions, the model does not account for feedbacks such as changes in global ocean circulation and ice sheet dynamics that, when investigated with a full Earth System model, would further refine our results. The challenge remains for modeling the inception and growth of the ice sheets.

REFERENCES AND NOTES

1. T. Westerhold *et al.*, An astronomically dated record of Earth's climate and its predictability over the last 66 million years. *Science* **369**, 1383–1387 (2020). doi: 10.1126/science.aba6853; pmid: 32913105
2. J. Zachos, M. Pagani, L. Sloan, E. Thomas, K. Billups, Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–693 (2001). doi: 10.1126/science.1059412; pmid: 11326091
3. N. J. Shackleton, N. D. Opdyke, in *GSA Memoirs*. R. M. Cline, J. D. Hays, Eds. (The Geological Society of America, 1976); vol. 145, pp. 449–464.
4. N. G. Pisias, T. C. Moore Jr., The evolution of Pleistocene climate: A time series approach. *Earth Planet. Sci. Lett.* **52**, 450–458 (1981). doi: 10.1016/0012-821X(81)90197-7
5. P. U. Clark *et al.*, The middle Pleistocene transition: Characteristics. mechanisms, and implications for long-term changes in atmospheric PCO_2. *Quat. Sci. Rev.* **25**, 3150–3184 (2006). doi: 10.1016/j.quascirev.2006.07.008
6. J. D. Hays, J. Imbrie, N. J. Shackleton, Variations in the Earth's orbit: Pacemaker of the ice ages. *Science* **194**, 1121–1132 (1976). doi: 10.1126/science.194.4270.1121; pmid: 17790893
7. J. Imbrie *et al.*, On the structure and origin of major glaciation cycles 1. Linear responses to Milankovitch forcing. *Paleoceanography* **7**, 701–738 (1992). doi: 10.1029/92PA02253
8. J. Imbrie *et al.*, On the structure and origin of major glaciation cycles. 2. The 100,000-year cycle. *Paleoceanography* **8**, 699–735 (1993). doi: 10.1029/93PA02751
9. J. Hansen, M. Sato, G. Russell, P. Kharecha, Climate sensitivity, sea level and atmospheric carbon dioxide. *Philos. Trans. A Math. Phys. Eng. Sci.* **371**, 20120294 (2013). doi: 10.1098/rsta.2012.0294; pmid: 24043864
10. N. G. Pisias *et al.*, High resolution stratigraphic correlation of benthic oxygen isotopic records spanning the last 300,000 years. *Mar. Geol.* **56**, 119–136 (1984). doi: 10.1016/0025-3227(84)90009-4
11. L. E. Lisiecki, M. E. Raymo, A Pliocene-Pleistocene stack of 57 globally distributed benthic δ18O records. *Paleoceanography* **20**, 2004PA001071 (2005). doi: 10.1016/0025-3227(84)90009-4
12. S. Ahn, D. Khider, L. E. Lisiecki, C. E. Lawrence, A probabilistic Pliocene–Pleistocene stack of benthic $\delta^{18}O$ using a profile hidden Markov model. *Dynamics and Statistics of the Climate System* **2**, 1–16 (2017). doi: 10.1093/climsys/dzx002
13. P. Huybers, Glacial variability over the last two million years: An extended depth-derived age model, continuous obliquity pacing, and the Pleistocene progression. *Quat. Sci. Rev.* **26**, 37–55 (2007). doi: 10.1016/j.quascirev.2006.07.013
14. P. Bajo *et al.*, Persistent influence of obliquity on ice age terminations since the Middle Pleistocene transition. *Science* **367**, 1235–1239 (2020). doi: 10.1126/science.aaw1114; pmid: 32165584
15. M. E. Raymo, R. Kozdon, D. Evans, L. Lisiecki, H. L. Ford, The accuracy of mid-Pliocene $\delta^{18}O$-based ice volume and sea level reconstructions. *Earth Sci. Rev.* **177**, 291–302 (2018). doi: 10.1016/j.earscirev.2017.11.022
16. D. P. Schrag, Effects of diagenesis on the isotopic record of late Paleogene tropical sea surface temperatures. *Chem. Geol.* **161**, 215–224 (1999). doi: 10.1016/S0009-2541(99)00088-1
17. H. J. Spero, J. Bijma, D. W. Lea, B. E. Bemis, Effect of seawater carbonate concentration on foraminiferal carbon and oxygen isotopes. *Nature* **390**, 497–500 (1997). doi: 10.1038/37333
18. P. U. Clark *et al.*, Mean ocean temperature change and decomposition of the benthic δ^{18O} record over the past 4.5 million years. *Clim. Past* **21**, 973–1000 (2025). doi: 10.5194/cp-21-973-2025
19. M. J. Winnick, J. K. Caves, Oxygen isotope mass-balance constraints on Pliocene sea level and East Antarctic Ice Sheet stability. *Geology* **43**, 879–882 (2015). doi: 10.1130/G36999.1
20. E. Gasson, R. M. DeConto, D. Pollard, Modeling the oxygen isotope composition of the Antarctic ice sheet and its significance to Pliocene sea level. *Geology* **44**, 827–830 (2016). doi: 10.1130/G38104.1
21. P. U. Clark, J. D. Shakun, Y. Rosenthal, P. Köhler, P. J. Bartlein, Global and regional temperature change over the past 4.5 million years. *Science* **383**, 884–890 (2024). doi: 10.1126/science.adi1908; pmid: 38386742
22. K. G. Miller *et al.*, Cenozoic sea-level and cryospheric evolution from deep-sea geochemical and continental margin records. *Sci. Adv.* **6**, eaaz1346 (2020). doi: 10.1126/sciadv.aaz1346; pmid: 32440543
23. J. D. Shakun, D. W. Lea, L. E. Lisiecki, M. E. Raymo, An 800-kyr record of global surface ocean delta O-18 and implications for ice volume-temperature coupling. *Earth Planet. Sci. Lett.* **426**, 58–68 (2015). doi: 10.1016/j.epsl.2015.05.042
24. C. Waelbroeck *et al.*, Sea-level and deep water temperature changes derived from benthic foraminifera isotopic records. *Quat. Sci. Rev.* **21**, 295–305 (2002). doi: 10.1016/S0277-3791(01)00101-9
25. S. L. Bates, M. Siddall, C. Waelbroeck, Hydrographic variations in deep ocean temperature over the mid-Pleistocene transition. *Quat. Sci. Rev.* **88**, 147–158 (2014). doi: 10.1016/j.quascirev.2014.01.020
26. K. A. Jakob *et al.*, A new sea-level record for the Neogene/Quaternary boundary reveals transition to a more stable East Antarctic Ice Sheet. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 30980–30987 (2020). doi: 10.1073/pnas.2004209117; pmid: 33229561
27. E. J. Rohling *et al.*, Comparison and synthesis of sea-level and deep-sea temperature variations over the past 40 million years. *Reviews of Geophysics* **60**, e2022RG000775 (2022). doi: 10.1029/2022RG000775
28. E. J. Rohling *et al.*, Sea level and deep-sea temperature reconstructions suggest quasi-stable states and critical transitions over the past 40 million years. *Sci. Adv.* **7**, eabf5326 (2021). doi: 10.1126/sciadv.abf5326; pmid: 34172440
29. P. R. Liautaud, D. A. Hodell, P. J. Huybers, Detection of significant climatic precession variability in early Pleistocene glacial cycles. *Earth Planet. Sci. Lett.* **536**, 116137 (2020). doi: 10.1016/j.epsl.2020.116137
30. S. Barker, L. E. Lisiecki, G. Knorr, S. Nuber, P. C. Tzedakis, Distinct roles for precession, obliquity, and eccentricity in Pleistocene 100-kyr glacial cycles. *Science* **387**, eadp3491 (2025). doi: 10.1126/science.adp3491; pmid: 40014707
31. P. Huybers, Early Pleistocene glacial cycles and the integrated summer insolation forcing. *Science* **313**, 508–511 (2006). doi: 10.1126/science.1125249; pmid: 16794041
32. M. A. Maslin, X. S. Li, M. F. Loutre, A. Berger, The contribution of orbital forcing to the progressive intensification of Northern Hemisphere glaciation. *Quat. Sci. Rev.* **17**, 411–426 (1998). doi: 10.1016/S0277-3791(97)00047-4
33. H. Cheng *et al.*, Ice age terminations. *Science* **326**, 248–252 (2009). doi: 10.1126/science.1177840; pmid: 19815769
34. H. Cheng *et al.*, The Asian monsoon over the past 640,000 years and ice age terminations. *Nature* **534**, 640–646 (2016). doi: 10.1038/nature18591; pmid: 27357793
35. P. C. Tzedakis, M. Crucifix, T. Mitsui, E. W. Wolff, A simple rule to determine which insolation cycles lead to interglacials. *Nature* **542**, 427–432 (2017). doi: 10.1038/nature21364; pmid: 28230118
36. P. Köhler, R. S. W. van de Wal, Interglacials of the Quaternary defined by northern hemispheric land ice distribution outside of Greenland. *Nat. Commun.* **11**, 5124 (2020). doi: 10.1038/s41467-020-18897-5; pmid: 33046715
37. O. A. Dumitru *et al.*, Constraints on global mean sea level during Pliocene warmth. *Nature* **574**, 233–236 (2019). doi: 10.1038/s41586-019-1543-2; pmid: 31471591
38. K. G. Miller *et al.*, High tide of the warm Pliocene: Implications of global sea level for Antarctic deglaciation. *Geology* **40**, 407–410 (2012). doi: 10.1130/G32869.1
39. F. D. Richards *et al.*, Geodynamically corrected Pliocene shoreline elevations in Australia consistent with midrange projections of Antarctic ice loss. *Sci. Adv.* **9**, eadg3035 (2023). doi: 10.1126/sciadv.adg3035; pmid: 37976352
40. G. S. Dwyer, M. A. Chandler, Mid-Pliocene sea level and continental ice volume based on coupled benthic Mg/Ca palaeotemperatures and oxygen isotopes. *Philos. Trans. A.* **367**, 157–168 (2009). doi: 10.1098/rsta.2008.0222; pmid: 18854304
41. S. Sosdian, Y. Rosenthal, Deep-sea temperature and ice volume changes across the Pliocene-Pleistocene climate transitions. *Science* **325**, 306–310 (2009). doi: 10.1126/science.1169938; pmid: 19608915
42. G. Balco, C. W. Rovey II, Absolute chronology for major Pleistocene advances of the Laurentide Ice Sheet. *Geology* **38**, 795–798 (2010). doi: 10.1130/G30946.1
43. A. J. Hidy, J. C. Gosse, D. G. Froese, J. D. Bond, D. H. Rood, A latest Pliocene age for the earliest and most extensive Cordilleran Ice Sheet in northwestern Canada. *Quat. Sci. Rev.* **61**, 77–84 (2013). doi: 10.1016/j.quascirev.2012.11.009

44. J. D. Shakun, M. E. Raymo, D. W. Lea, An early Pleistocene Mg/Ca-$\delta^{18}O$ record from the Gulf of Mexico: Evaluating ice sheet size and pacing in the 41-kyr world. *Paleoceanography* **31**, 1011–1027 (2016). doi: 10.1002/2016PA002956
45. K. Blake-Mizen *et al.*, Southern Greenland glaciation and Western Boundary Undercurrent evolution recorded on Eirik Drift during the late Pliocene intensification of Northern Hemisphere glaciation. *Quat. Sci. Rev.* **209**, 40–51 (2019). doi: 10.1016/j.quascirev.2019.01.015
46. B. R. Rea *et al.*, Extensive marine-terminating ice sheets in Europe from 2.5 million years ago. *Sci. Adv.* **4**, eaar8327 (2018). doi: 10.1126/sciadv.aar8327; pmid: 29928693
47. E. Jansen, T. Fronval, F. Rack, J. E. T. Channell, Pliocene-Pleistocene ice rafting history and cyclicity in the Nordic Seas during the last 3.5 Myr. *Paleoceanography* **15**, 709–721 (2000). doi: 10.1029/1999PA000435
48. G. H. Haug, D. M. Sigman, R. Tiedemann, T. F. Pedersen, M. Sarnthein, Onset of permanent stratification in the subarctic Pacific Ocean. *Nature* **401**, 779–782 (1999). doi: 10.1038/44550
49. N. J. Shackleton, N. D. Opdyke, Oxygen isotope and paleomagnetic evidence for early Northern Hemisphere glaciation. *Nature* **270**, 216–219 (1977). doi: 10.1038/270216a0
50. M. E. Raymo, The initiation of Northern Hemisphere glaciation. *Annu. Rev. Earth Planet. Sci.* **22**, 353–383 (1994). doi: 10.1146/annurev.ea.22.050194.002033
51. P. U. Clark, D. Pollard, Origin of the middle Pleistocene transition by ice sheet erosion of regolith. *Paleoceanography* **13**, 1–9 (1998). doi: 10.1029/97PA02660
52. M. E. Raymo, K. Nisancioglu, The 41 kyr world: Milankovitch's other unsolved mystery. *Paleoceanography* **18**, 2002PA000791 (2003). doi: 10.1029/97PA02660
53. P. Huybers, E. Tziperman, Integrated summer insolation forcing and 40,000-year glacial cycles: The perspective from an ice sheet/energy-balance model. *Paleoceanography* **23**, PA1208 (2008). doi: 10.1029/2007PA001463
54. C. R. Tabor, C. J. Poulsen, D. Pollard, How obliquity cycles powered early Pleistocene global ice-volume variability. *Geophys. Res. Lett.* **42**, 1871–1879 (2015). doi: 10.1002/2015GL063322
55. J. Boellstorff, North American Pleistocene stages reconsidered in light of probable Pliocene-Pleistocene continental glaciation. *Science* **202**, 305–307 (1978). doi: 10.1126/science.202.4365.305; pmid: 17817644
56. M. E. Raymo, L. E. Lisiecki, K. H. Nisancioglu, Plio-Pleistocene ice volume, Antarctic climate, and the global $\delta_{18}O$ record. *Science* **313**, 492–495 (2006). doi: 10.1126/science.1123296; pmid: 16794038
57. K. A. Dyez, B. Hönisch, G. A. Schmidt, Early Pleistocene obliquity-scale pCO_2 variability at ~1.5 million years ago. *Paleoceanogr. Paleoclimatol.* **33**, 1270–1291 (2018). doi: 10.1029/2018PA003349; pmid: 32715282
58. Z. Liu, H. Yang, Extratropical control of tropical climate, the atmospheric bridge and oceanic tunnel. *Geophys. Res. Lett.* **30**, 2002GL016492 (2003). doi: 10.1029/2002GL016492
59. S. Y. Lee, C. J. Poulsen, Sea ice control of Plio-Pleistocene tropical Pacific climate evolution. *Earth Planet. Sci. Lett.* **248**, 253–262 (2006). doi: 10.1016/j.epsl.2006.05.030
60. T. D. Herbert, L. C. Peterson, K. T. Lawrence, Z. Liu, Tropical ocean temperatures over the past 3.5 million years. *Science* **328**, 1530–1534 (2010). doi: 10.1126/science.1185435; pmid: 20558711
61. P. Huybers, P. Molnar, Tropical cooling and the onset of North American glaciation. *Clim. Past* **3**, 549–557 (2007). doi: 10.5194/cp-3-549-2007
62. P. U. Clark, S. W. Hostetler, N. G. Pisias, A. Schmittner, K. J. Meissner, in *Ocean Circulation: Mechanisms and Impacts*. A. Schmittner, J. Chiang, S. Hemming, Eds. (American Geophysical Union, 2007); pp. 209–246.
63. M. P. Erb, C. S. Jackson, A. J. Broccoli, Using Single-Forcing GCM Simulations to Reconstruct and Interpret Quaternary Climate Change. *J. Clim.* **28**, 9746–9767 (2015). doi: 10.1175/JCLI-D-15-0329.1
64. J. Fyke, O. Sergienko, M. Lofverstrom, S. Price, J. T. M. Lenaerts, An overview of interactions and feedbacks between ice sheets and the Earth system. *Rev. Geophys.* **56**, 361–408 (2018). doi: 10.1029/2018RG000600
65. X. Li *et al.*, Mid-Pliocene westerlies from PlioMIP simulations. *Adv. Atmos. Sci.* **32**, 909–923 (2015). doi: 10.1007/s00376-014-4171-7
66. S. Hofer, A. J. Tedstone, X. Fettweis, J. L. Bamber, Decreasing cloud cover drives the recent mass loss on the Greenland Ice Sheet. *Sci. Adv.* **3**, e1700584 (2017). doi: 10.1126/sciadv.1700584; pmid: 28782014
67. N. Sagoo *et al.*, Observationally constrained cloud phase unmasks orbitally driven climate feedbacks. *Geophysical Research Letters* **48**, e2020GL091873 (2021). doi: 10.1126/sciadv.1700584; pmid: 28782014
68. N. Unger, X. Yue, Strong chemistry- climate feedbacks in the Pliocene. *Geophys. Res. Lett.* **41**, 527–533 (2014). doi: 10.1002/2013GL058773
69. D. Pollard, S. L. Thompson, Climate and ice sheet mass balance at the Last Glacial Maximum from GENESIS version 2 global climate model. *Quat. Sci. Rev.* **16**, 841–863 (1997). doi: 10.1016/S0277-3791(96)00115-1
70. S. L. Thompson, D. Pollard, Greenland and Antarctic mass balances for present and doubled atmospheric CO_2 from the GENESIS version-2 global climate model. *J. Clim.* **10**, 871–900 (1997). https://doi.org/10.1175/1520-0442(1997)010<0871:GAAMBF>2.0.CO;2
71. J. Laskar *et al.*, A long term numerical solution for the insolation quantities of the Earth. *Astron. Astrophys.* **428**, 261–285 (2004). doi: 10.1051/0004-6361:20041335
72. P. Köhler, Atmospheric CO_2 concentration based on boron isotopes versus simulations of the global carbon cycle during the Plio-Pleistocene. *Paleoceanography and Paleoclimatology* **38**, e2022PA004439 (2023). doi: 10.1051/0004-6361:20041335
73. A. Berger, X. S. Li, M. F. Loutre, Modelling northern hemisphere ice volume over the last 3 Ma. *Quat. Sci. Rev.* **18**, 1–11 (1999). doi: 10.1016/S0277-3791(98)00033-X
74. M. Willeit, A. Ganopolski, R. Calov, V. Brovkin, Mid-Pleistocene transition in glacial cycles explained by declining CO_2 and regolith removal. *Sci. Adv.* **5**, eaav7337 (2019). doi: 10.1126/sciadv.aav7337; pmid: 30949580
75. E. Tziperman, H. Gildor, On the mid-Pleistocene transition to 100-kyr glacial cycles and the asymmetry between glaciation and deglaciation times. *Paleoceanography* **18**, PA000627 (2003). doi: 10.1029/2001pa000627
76. P. Huybers, C. Wunsch, Obliquity pacing of the late Pleistocene glacial terminations. *Nature* **434**, 491–494 (2005). doi: 10.1038/nature03401; pmid: 15791252
77. T. B. Chalk *et al.*, Causes of ice age intensification across the Mid-Pleistocene Transition. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 13114–13119 (2017). doi: 10.1073/pnas.1702143114; pmid: 29180424
78. N. J. Shackleton, The 100,000-year Ice-Age cycle identified and found to lag temperature, carbon dioxide, and orbital eccentricity. *Science* **289**, 1897–1902 (2000). doi: 10.1126/science.289.5486.1897; pmid: 10988063
79. J. Imbrie, J. Z. Imbrie, Modeling the climatic response to orbital variations. *Science* **207**, 943–953 (1980). doi: 10.1126/science.207.4434.943; pmid: 17830447
80. A. Abe-Ouchi *et al.*, Insolation-driven 100,000-year glacial cycles and hysteresis of ice sheet volume. *Nature* **500**, 190–193 (2013). doi: 10.1038/nature12374; pmid: 23925242
81. D. Pollard, A simple ice sheet model yields realistic 100 kyr glacial cycles. *Nature* **296**, 334–338 (1982). doi: 10.1038/296334a0
82. P. U. Clark, R. B. Alley, D. Pollard, Northern Hemisphere ice sheet influences on global climate change. *Science* **286**, 1104–1111 (1999). doi: 10.1126/science.286.5442.1104
83. W. F. Ruddiman, M. E. Raymo, D. G. Martinson, B. M. Clement, J. Backman, Pleistocene evolution: Northern hemisphere ice sheets and North Atlantic ocean. *Paleoceanography* **4**, 353–412 (1989). doi: 10.1029/PA004i004p00353
84. D. M. Sigman *et al.*, The Southern Ocean during the ice ages: A review of the Antarctic surface isolation hypothesis, with comparison to the North Pacific. *Quat. Sci. Rev.* **254**, 106732 (2021). doi: 10.1016/j.quascirev.2020.106732
85. A. Martínez-Garcia *et al.*, Southern Ocean dust-climate coupling over the past four million years. *Nature* **476**, 312–315 (2011). doi: 10.1038/nature10310; pmid: 21814203
86. A. P. Hasenfratz *et al.*, The residence time of Southern Ocean surface waters and the 100,000-year ice age cycle. *Science* **363**, 1080–1084 (2019). doi: 10.1126/science.aat7067; pmid: 30846597
87. R. Ferrari *et al.*, Antarctic sea ice control on ocean circulation in present and glacial climates. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8753–8758 (2014). doi: 10.1073/pnas.1323922111; pmid: 24889624
88. M. F. Jansen, Glacial ocean circulation and stratification explained by reduced atmospheric temperature. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 45–50 (2017). doi: 10.1073/pnas.1610438113; pmid: 27994158
89. A. Marzocchi, M. F. Jansen, Global cooling linked to increased glacial carbon storage via changes in Antarctic sea ice. *Nat. Geosci.* **12**, 1001–1005 (2019). doi: 10.1038/s41561-019-0466-8
90. P. Köhler, R. Bintanja, The carbon cycle during the Mid Pleistocene Transition: The Southern Ocean decoupling hypothesis. *Clim. Past* **4**, 311–332 (2008). doi: 10.5194/cp-4-311-2008
91. D. Pollard, A coupled climate-ice sheet model applied to the Quaternary ice ages. *J. Geophys. Res.* **88**, 7705–7718 (1983). doi: 10.1029/JC088iC12p07705
92. S. J. Marshall, P. U. Clark, Basal temperature evolution of North American ice sheets and implications for the 100-kyr cycle. *Geophys. Res. Lett.* **29**, GL015192 (2002). doi: 10.1029/2002GL015192
93. P. U. Clark *et al.*, Oceanic forcing of penultimate deglacial and last interglacial sea level rise. *Nature* **577**, 660–664 (2020). doi: 10.1038/s41586-020-1931-7; pmid: 31996820
94. M. E. Raymo, The timing of major climate terminations. *Paleoceanography* **12**, 577–585 (1997). doi: 10.1029/97PA01169
95. W. S. Broecker, J. van Donk, Insolation changes, ice volumes, and the O_{18} record in deep-sea cores. *Rev. Geophys.* **8**, 169–198 (1970). doi: 10.1029/RG008i001p00169
96. N. J. Shackleton, The oxygen isotope stratigraphic record of the Late Pleistocene. *Philos. Trans. R. Soc. Lond.* **280B**, 169–182 (1977).
97. S. Barker *et al.*, Persistent influence of precession on northern ice sheet variability since the early Pleistocene. *Science* **376**, 961–967 (2022). doi: 10.1126/science.abm4033; pmid: 35617392
98. M. Siddall, B. Hönisch, C. Waelbroeck, P. Huybers, Changes in deep Pacific temperature during the mid-Pleistocene transition and Quaternary. *Quat. Sci. Rev.* **29**, 170–181 (2010). doi: 10.1016/j.quascirev.2009.05.011
99. R. M. Spratt, L. E. Lisiecki, A Late Pleistocene sea level stack. *Clim. Past* **12**, 1079–1092 (2016). doi: 10.5194/cp-12-1079-2016
100. J. E. Hansen *et al.*, Global warming in the pipeline. *Oxford Open Climate Change* **3**, kgad008 (2023). doi: 10.5194/cp-12-1079-2016
101. H. Elderfield *et al.*, Evolution of ocean temperature and ice volume through the mid-Pleistocene climate transition. *Science* **337**, 704–709 (2012). doi: 10.1126/science.1221294; pmid: 22879512

102. S. C. Woodard *et al.*, Antarctic role in Northern Hemisphere glaciation. *Science* **346**, 847–851 (2014). doi: 10.1126/science.1255586; pmid: 25342658
103. H. L. Ford, M. E. Raymo, Regional and global signals in seawater $\delta^{18}O$ records across the mid-Pleistocene transition. *Geology* **48**, 113–117 (2020). doi: 10.1130/G46546.1
104. A. C. Mix, W. F. Ruddiman, Oxygen-isotope analyses and Pleistocene ice volumes. *Quat. Res.* **21**, 1–20 (1984). doi: 10.1016/0033-5894(84)90085-1
105. G. R. Grant *et al.*, Reduced magnitude of Early Pleistocene intensification of Northern Hemisphere Glaciation. *Quat. Sci. Rev.* **349**, 109096 (2025). doi: 10.1016/j.quascirev.2024.109096
106. G. R. Grant *et al.*, The amplitude and origin of sea level variability during the Pliocene epoch. *Nature* **574**, 237–241 (2019). doi: 10.1038/s41586-019-1619-z; pmid: 31578526
107. D. P. Schrag *et al.*, The oxygen isotopic composition of seawater during the Last Glacial Maximum. *Quat. Sci. Rev.* **21**, 331–342 (2002). doi: 10.1016/S0277-3791(01)00110-X
108. K. Lambeck, H. Rouby, A. Purcell, Y. Sun, M. Sambridge, Sea level and global ice volumes from the Last Glacial Maximum to the Holocene. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15296–15303 (2014). doi: 10.1073/pnas.1411762111; pmid: 25313072
109. M. B. Priestley, Evolutionary spectra and non-stationary processes. *J. R. Stat. Soc. Series B Stat. Methodol.* **27**, 204–229 (1965). doi: 10.1111/j.2517-6161.1965.tb01488.x
110. S. R. Meyers, astrochron: A Computational Tool for Astrochronology. (2025); https://cran.r-project.org/package=astrochron
111. R Core Team, R: The R Project for Statistical Computing (R Foundation for Statistical Computing, 2024); www.R-project.org/
112. D. J. Thomson, Spectrum estimation and harmonic analysis. *Proc. IEEE* **70**, 1055–1096 (1982). doi: 10.1109/PROC.1982.12433
113. U. Ligges, T. Short, P. Kienzle, signal: Signal Processing. R package version 1.5 (2024); https://CRAN.R-project.org/package=signal
114. P. Howell, N. Pisias, J. Ballance, J. Baughman, L. Ochs, ARAND Time-Series Analysis Software (Brown University, 2006).
115. I. Janssens, P. Huybrechts, The Treatment of Meltwater Retardation in Mass-Balance Parameterizations of the Greenland Ice Sheet. *Ann. Glaciol.* **31**, 133–140 (2000). doi: 10.3189/172756400781819941
116. B. Hönisch, N. G. Hemming, D. Archer, M. Siddall, J. F. McManus, Atmospheric carbon dioxide concentration across the mid-Pleistocene transition. *Science* **324**, 1551–1554 (2009). doi: 10.1126/science.1171477; pmid: 19541994
117. G. Bartoli, B. Hönisch, R. E. Zeebe, Atmospheric CO_2 decline during the Pliocene intensification of Northern Hemisphere glaciations. *Paleoceanography* **26**, 2010PA002055 (2011). doi: 10.1029/2010PA002055
118. L. Tarasov, A. S. Dyke, R. M. Neal, W. R. Peltier, A data-calibrated distribution of deglacial chronologies for the North American ice complex from glaciological modeling. *Earth Planet. Sci. Lett.* **315**, 30–40 (2012). doi: 10.1016/j.epsl.2011.09.010
119. O. A. Dumitru *et al.*, Sea level stands from the Western Mediterranean over the past 6.5 million years. *Sci. Rep.* **11**, 6681 (2021). doi: 10.1038/s41598-021-85719-z; pmid: 33731822
120. A. Dutton, J. M. Webster, D. Zwartz, K. Lambeck, B. Wohlfarth, Tropical tales of polar ice: Evidence of Last Interglacial polar ice sheet retreat recorded by fossil reefs of the granitic Seychelles islands. *Quat. Sci. Rev.* **107**, 182–196 (2015). doi: 10.1016/j.quascirev.2014.10.025
121. D. Ottesen, C. L. Batchelor, H. Løseth, H. Brunstad, 3D seismic evidence for a single Early Pleistocene glaciation of the central North Sea. *Sci. Adv.* **10**, eadq6089 (2024). doi: 10.1126/sciadv.adq6089; pmid: 39671491
122. B. Bereiter *et al.*, Revision of the EPICA Dome C CO_2 record from 800 to 600kyr before present. *Geophys. Res. Lett.* **42**, 542–549 (2015). doi: 10.1002/2014GL061957
123. M. Bouchet *et al.*, The Antarctic Ice Core Chronology 2023 (AICC2023) chronological framework and associated timescale for the European Project for Ice Coring in Antarctica (EPICA) Dome C ice core. *Clim. Past* **19**, 2257–2286 (2023). doi: 10.5194/cp-19-2257-2023
124. P. Huybers, Combined obliquity and precession pacing of late Pleistocene deglaciations. *Nature* **480**, 229–232 (2011). doi: 10.1038/nature10626; pmid: 22158246

ACKNOWLEDGMENTS

This publication contributed to Beyond EPICA, a project of the European Union's Horizon 2020 research and innovation program (Oldest Ice Core). This is Beyond EPICA publication number 45. **Funding:** This work was funded by the following: National Science Foundation OPP-2103032 (to P.U.C.); University of Oregon Fund for Faculty Excellence (to P.B.); National Science Foundation OCE-1810681 (to Z.L.); National Science Foundation OCE-1834208 (to Y.R.). **Author contributions:** Conceptualization: P.U.C. and J.D.S.; Methodology: J.D.S., Y.R., D.P., S.W.H., P.U.C., P.J.B., and N.G.P.; Investigation: P.U.C., J.D.S., Y.R., D.P., S.W.H., P.K., and P.J.B.; Writing – original draft: P.U.C.; Writing – review & editing: P.U.C., J.D.S., Y.R., D.P., S.W.H., P.K., P.J.B., J.M.G., D.P.S., C.Z., Z.L., and N.G.P. **Competing interests:** Authors declare that they have no competing interests. **Data Availability Statement:** All data are available in the SM.
License information: Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/content/page/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adv8389
Figs. S1 to S10; Table S1; References (*125*–*131*); Equations and parameters to convert $\delta^{18}O_{sw}$ to sea level

Submitted 9 January 2025; accepted 25 August 2025

10.1126/science.adv8389

ORGANIC CHEMISTRY

The total synthesis of (−)-spiroaspertrione A: A divinylcyclopropane rearrangement approach

Wenbo Huang[1,2]†, Lu Pan[1,2]†, Heng Zhao[3], Fabian Schneider[1]*, Tanja Gaich[2]*

The rise of multidrug-resistant pathogens poses a major threat to global health, with methicillin-resistant *Staphylococcus aureus* (MRSA) among the most challenging. One promising approach to overcoming resistance is using small molecules that resensitize MRSA to existing drugs. Here, we report the enantioselective total synthesis of one such promising candidate, (−)-spiroaspertrione A, a complex meroterpenoid of the andiconin family. This natural product has long eluded synthesis because of its densely functionalized polycyclic backbone. Our route features a stereoselective Diels-Alder cycloaddition, followed by a key divinylcyclopropane rearrangement forming the spirobicyclo[3.2.2]nonane core, which proved to be reversible and was further investigated by density functional theory calculations. Strategic late-stage functionalization of the compact cage architecture enabled access to the natural product and provided evidence for a plausible biosynthetic relationship with (−)-aspermerodione.

The structural diversity and complexity of fungal meroterpenoids arise primarily from the catalytic versatility of various enzymes that orchestrate reactions with distinct patterns and stereospecificity at multiple stages of biosynthesis, leading to pathway divergence. Since the isolation of the first 3,5-dimethylorsellinic acid (DMOA)–derived meroterpenoid, andibenin B (**1**) (*1*), in 1976 from *Aspergillus variecolor*, >200 secondary metabolites have been identified in this category (*2*–*4*). These compounds share a common polyketide origin but diverge considerably through distinct terpene coupling and tailoring reactions. Among them, a defined subgroup features andiconin (**2**) (*5*), which serves as a gateway compound to a series of closely related yet architecturally diverse natural products (Fig. 1A, **1** to **6**). One prominent example is (−)-spiroaspertrione A (**3**) (*6*), first isolated in 2017 from a culture of *Aspergillus* sp. TJ23. Subsequent biological evaluations revealed that (−)-spiroaspertrione A (**3**) exhibits potent activity against methicillin-resistant *Staphylococcus aureus* (MRSA), with a minimum inhibitory concentration (MIC) value of 4 μg/ml. Additionally, (−)-spiroaspertrione A (**3**) resensitizes MRSA to oxacillin, reducing its MIC value against MRSA from 32 to 1 μg/ml (*6*). Despite this marked bioactivity, further in vivo evaluation and mechanistic studies were hindered by the limited amount of material available from the fungus *Aspergillus*. This scarcity (10.5 mg of **3** isolated from 25 liters of culture), combined with the lack of synthetic access to the andiconin-derived meroterpenoids, underscores the value of (−)-spiroaspertrione A (**3**) as a synthetic target.

[1]Shanghai Research Center for Bioactive Molecule Synthesis (BiMoS), Shanghai University of Traditional Chinese Medicine, Shanghai, China. [2] Department of Chemistry, University of Konstanz, Konstanz, Germany. [3] Department of Chemistry, University of Fribourg, Fribourg, Switzerland. ***Corresponding author. Email: fschneider@shutcm.edu.cn (F.S.); tanja.gaich@uni-konstanz.de (T.G.)** †These authors contributed equally to this work.

In contrast to the more common tricyclo[4.3.1.0^{3,7}]decane and tricyclo[4.3.1.0^{3,8}]decane frameworks typically found in andiconin-derived meroterpenoids (Fig. 1A, green and blue, respectively), (−)-spiroaspertrione A (**3**) features a spirobicyclo[3.2.2]nonane core (red), which presents several synthetic challenges. Foremost among these is the construction of the spirobicyclic scaffold bearing adjacent quaternary centers at C8 and C2′, which requires overcoming substantial steric hindrance. Furthermore, functionalization of the core is impeded by stereoelectronic interactions along the cage periphery and by the presence of the vicinal C9 ketone moiety. Another congener, (−)-aspermerodione (**4**) (*7*), which shares the same carbon skeleton as (−)-spiroaspertrione A (**3**), also captured our attention. Despite its distinct molecular geometry, (−)-aspermerodione (**4**) exhibits comparable biological activity to (−)-spiroaspertrione A (**3**) (*7*), albeit with slightly reduced potency, suggesting the involvement of a common bioactive intermediate and a shared mode of action, as well as the possibility of interconversion. This hypothesis is supported by Matsuda and Abe's proposal (*8*), that (−)-aspermerodione (**4**) is a downstream biosynthetic intermediate of (−)-spiroaspertrione A (**3**) (Fig. 1B) that is formed through hydrolysis and methylation. Inspired by this insight, our retrosynthetic analysis envisioned accessing (−)-spiroaspertrione A (**3**) through a transannular elimination of methanol from (−)-aspermerodione (**4**) (blue arrows). Our synthetic route to (−)-aspermerodione (**4**) used a divinylcyclopropane rearrangement (DVCPR) (*9*, *10*) of **(*E*)-8** to assemble the spirobicyclo[3.2.2]nonane core in **7**. The precursor **(*E*)-8** was delineated from diene **9**, wherein the C8 spiro center was constructed stereoselectively through an intermolecular Diels-Alder cycloaddition between diene **10** and enone **11**.

Route to the divinylcyclopropane intermediate

Our synthesis commenced with (+)-enoxolone (**12**), an inexpensive (0.90 euros/g) chiral pool building block. After Jones oxidation of **12** to ketone **13**, flash vacuum pyrolysis triggered a *retro*-Diels-Alder reaction that directly afforded bicyclic building block **14** (*11*), which was converted to the literature-known enone **11** through four functional group interconversions. A Diels-Alder cycloaddition between enone **11** and diene **10** established the C8 spiro center. The best results were obtained using Jung's mixed Lewis acid catalysis [Al(CH$_3$)$_3$–AlBr$_3$] (*12*), followed by oxidation with 2,3-dichloro-5,6-dicyano-1,4-benzoquinone (DDQ) (*13*), yielding spiroenone **15** with excellent diastereomeric

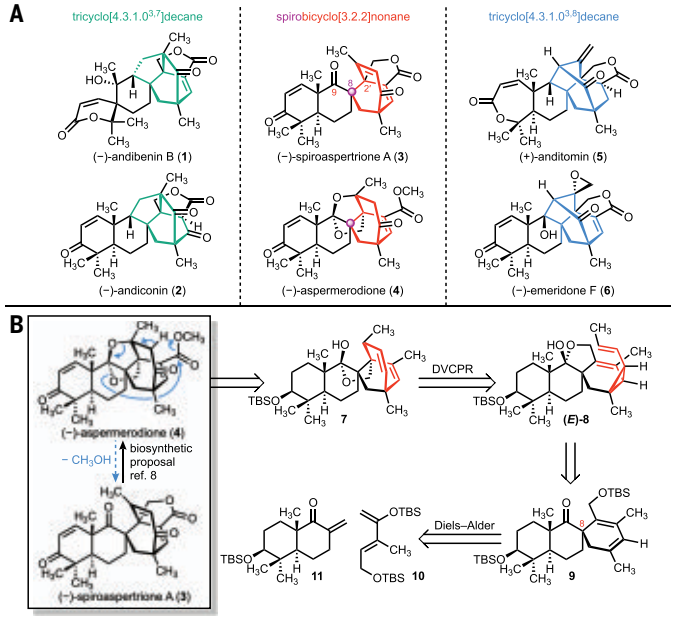


Fig. 1. Andiconin-derived meroterpenoids. (**A**) Representative natural products. (**B**) Retrosynthetic analysis of (−)-spiroaspertrione A (**3**). TBS, *tert*-butyldimethylsilyl.

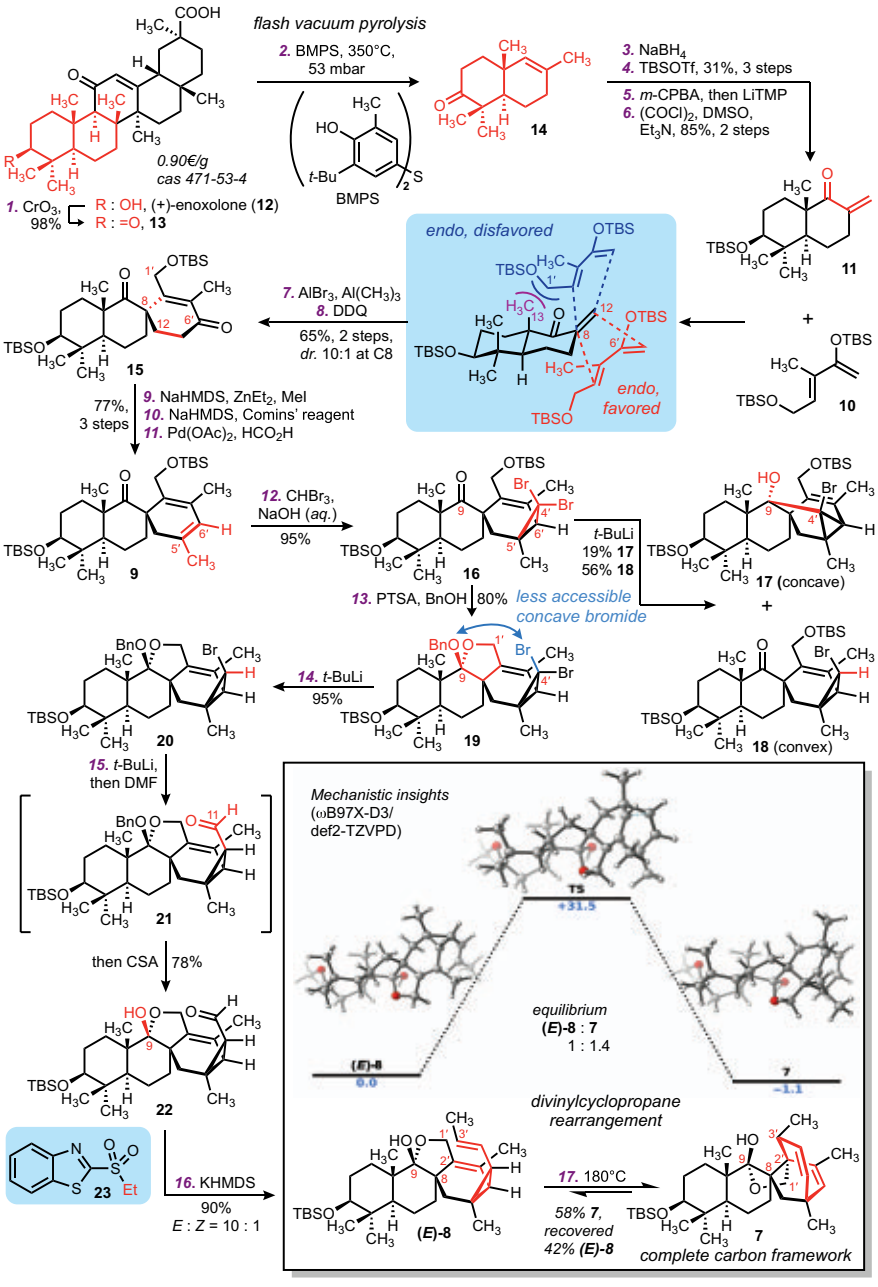


Fig. 2. Construction of the carbon framework through DVCPR. Reagents and conditions: (1) CrO_3, acetone, 0°C; (2) 4,4′-thiobis(2-methyl-6-*tert*-butylphenol) (BMPS), 350°C, 53 mbar; (3) $NaBH_4$, tetrahydrofuran (THF)–CH_3OH (3:1), 0°C; (4) *tert*-butyldimethylsilyl triflate (TBSOTf), triethylamine (Et_3N), CH_2Cl_2, 0°C; (5) *meta*-chloroperoxybenzoic acid (*m*-CPBA), $NaHCO_3$, 0°C, then lithium tetramethylpiperidide (LiTMP), THF, 0°C; (6) $(COCl)_2$, DMSO, Et_3N, −78°C to room temperature (RT); (7) $Al(CH_3)_3$, $AlBr_3$, CH_2Cl_2, −15°C, diastereomeric ratio 10:1 at C8); (8) DDQ, lutidine, toluene, RT; (9) sodium bis(trimethylsilyl)amide (NaHMDS), THF, 0°C, then diethylzinc ($ZnEt_2$), 1,3-dimethyl-3,4,5,6-tetrahydro-2(1*H*)-pyrimidinone (DMPU), CH_3I; (10) NaHMDS, THF, 0°C, then *N*-(5-chloropyridin-2-yl)-1,1,1-trifluoro-*N*-((trifluoromethyl)sulfonyl)methanesulfonamide (Comins' reagent); (11) palladium(II) acetate $Pd(OAc)_2$, triphenylphosphine (PPh_3), Et_3N, HCO_2H, THF, 70°C; (12) $CHBr_3$, benzyltriethylammonium chloride (TEBAC), NaOH (50 wt% aq.), RT; (13) PTSA, benzyl alcohol (BnOH), 4Å molecular sieves, toluene, RT; (14) *tert*-butyllithium (*t*-BuLi), THF, −78°C; (15) *t*-BuLi, −78 to 0°C, then −78°C, DMF, then CSA, RT; (16) 2-ethylsulfonyl-1,3-benzothiazole (**23**), potassium bis(trimethylsilyl)amide (KHMDS), THF, −78°C; and (17) toluene, 180°C. Mechanistic insights into the DVCPR are from DFT calculations. Energies (in kcal/mol) are relative to that of ***(E)*-8**, and all structures were calculated at the (ωB97X-D3/def2-TZVPD||BP86-D3BJ/def2-SVP) level of DFT.

control at the C8 spiro center (diastereomeric ratio, 10:1). This stereoselectivity was attributed to steric interactions between the C13 angular methyl group (Fig. 2, first blue box, highlighted in purple) and diene **10** in the transition state, as illustrated by two distinct orientations (highlighted in blue and red). Proceeding from enone **15**, α-methylation of the C6′ ketone was facilitated by $ZnEt_2$ (*14*, *15*), which minimized the basicity of the enolate and thus suppressed undesired α-dimethylation. Subsequent vinyl triflate formation using Comins' reagent followed by reductive detriflation yielded diene **9**. Next, various approaches were explored to construct divinylcyclopropane, including the classic combination of diazo compounds with transition metal catalysis (e.g., rhodium or copper) (*16*). However, the low reactivity of the C5′–C6′ trisubstituted olefin in **9** thwarted these attempts, resulting in recovery of the starting material. Nevertheless, cyclopropanation of **9** using dibromocarbene (*17*, *18*) provided dibromocyclopropane **16** in 95% yield as a single diastereomer.

Stereoselective lithium-bromine exchange on the gem-dibromocyclopropane **16** afforded the desired monobromide **18** in a 3:1 ratio, with tertiary alcohol **17** resulting from intramolecular addition of the C4′ concave carbanion to the C9 ketone. Extensive studies of this lithium-bromine exchange revealed that the C4′ cyclopropyl carbanion was configurationally stable, thus suggesting that the 3:1 ratio of bromide **18** and alcohol **17** reflects the kinetic rates of the lithium-bromide exchange between the convex and concave bromide substituents. To suppress the unwanted aldol addition, the C9 carbonyl in **16** was protected as acetal **19**, which also exerted concave or convex kinetic differentiation through steric interactions. Indeed, treatment of **19** with *tert*-butyllithium furnished the convex debromination product **20** as a single diastereomer. Further treatment of **20** with *tert*-butyllithium, followed by dimethylformamide (DMF) and camphorsulfonic acid (CSA), sequentially generated aldehyde **21** and hydrolyzed the C9 ketal in situ to afford hemiketal **22**. A modified Julia olefination (*19*) between aldehyde **22** and benzothiazole sulfone **23** provided ***(E)*-8** in 90% yield as a 10:1 mixture of (*E*) and (*Z*) isomers. Hydrolysis of the C9 benzyl ketal before the olefination was crucial for achieving high *E*/*Z* selectivity, because direct olefination of **21** under identical conditions gave poor stereoselectivity (*E*:*Z* = 1.4:1).

DVCPR and density functional theory calculations

Heating of ***(E)*-8** at 180°C established an equilibrium between it and the desired DVCPR product, **7** (1:1.4). This intermediate contains the complete carbon framework of (−)-spiroaspertrione A (***3***), comprising six contiguous stereogenic centers including three quaternary centers. Typically, DVCPR is an irreversible process because of the significant energy difference between the starting material and the product. However, in the

case of (***E***)-**8**, the energy released upon cyclopropane ring opening (~27 kcal/mol) is mitigated by the formation of the two adjacent quaternary centers C8 and C2′ in **7**, resulting in a marginal energy difference between (***E***)-**8** and **7**. Density functional theory (DFT) calculations (wB97X-D3/def2-TZVPD||BP86-D3BJ/def-SVP; see the supplementary materials) were used to study this equilibrium. The activation energy required for this transformation was determined to be 31.5 kcal/mol, consistent with the high experimental temperature (180°C), whereas the energy difference between **7** and (***E***)-**8** was only −1.1 kcal/mol, leading to the observed equilibrium.

Late-stage functionalization: Completion of the synthesis

Upon obtaining the DVCPR product **7**, our final objective was to establish the oxidation patterns at C3′, C4′, and C8′ of the bicyclic core and C1, C2, and C3 at the decalin moiety (Fig. 3A). However, attempts to introduce oxidation or desaturation at the C3′ position of **7** proved to be challenging. Several strategies were explored to address this hurdle. One approach leveraged the spatial proximity of the C9 hemiketal and C3′–H, as evidenced by nuclear Overhauser effect (NOE) studies and the formation of the bicyclic ketal in (−)-aspermerodione (**4**). Heating hemiketal **7** with lead tetraacetate in benzene at reflux temperature (*20*, *21*) achieved the desired intramolecular hydrogen abstraction, yielding bicyclic ketal **24**. However, the formation of the C3′ oxygen bridge deactivated the C11–C4′ double bond, impeding further progress. Treatment of ketal **24** with Schlosser's base (*n*-BuLi–KO*t*-Bu) (*22*–*24*) induced deprotonation at the C8′ allylic position, triggering a substantial skeletal rearrangement (Fig. 3A, blue arrows) that afforded barbaralane **25** quantitatively. However, attempts to convert alkene **25** into allylic alcohol **26** through Mukaiyama hydration (*25*) led to decomposition of the material.

In light of this observation, we prioritized the oxidation of the C11–C4′ double bond and aimed to desaturate the C3′ position using a Saegusa-Ito oxidation (*26*) of the C4′ ketone (Fig. 3B). Alkene **7** underwent allylic oxidation with selenium dioxide to the allylic alcohol **27**, which could then be further oxidized to lactone **28** quantitatively with manganese dioxide. The introduction of the C4′ ketone was achieved by a sequence comprising epoxidation of the C11–C4′ double bond of **28** with dimethyldioxirane (DMDO), titanium(III)–mediated reductive epoxide opening, and oxidation of the C4′ alcohol under Ley's conditions (*27*). Unfortunately, attempts to desaturate C3′–C11 within **31** using 2-iodoxybenzoic acid (IBX) or DDQ were unsuccessful. The C4′ ketone could be transformed into the corresponding silyl enol ether but failed to undergo further oxidation (Fig. 3B). In addition to the expected alcohol **30**, an unorthodox by-product, cyclobutanol **29**, bearing a new tricyclo[4.2.1.$0^{3,8}$]nonane core, was formed from the reductive epoxide opening by an intramolecular radical Michael addition (from C11 to C6′). This observation illustrates the close proximity of the cage periphery, which repeatedly led to unexpected chemical behavior of the cage system.

To circumvent these issues, we focused on generating the C11–C4′ epoxide with the intent to convert it into an allylic alcohol through the intermediacy of a selenohydrin (Fig. 4A). Treatment of alkene **7** with DMDO provided a diastereomeric mixture of epoxides ***32*** and ***33***. The subsequent reaction with lithium phenylselenide (PhSeLi) not only produced the expected selenohydrins ***34*-Se** and ***35*-Se**, but also directly yielded allylic alcohol **26** in moderate yield. Analyzing the results and conditions led to the conclusion that the desired allylic alcohol **26** was generated from epoxide ***33***, with PhSeLi acting primarily as a bulky base in this transformation.

The reaction was subsequently optimized to increase the selectivity for the formation of ***33***. After extensive screening (seen in Fig. 4, table 1), we determined that protic solvents capable of hydrogen bonding, such as methanol, increased the amount of ***33***. The origin of this diastereoselectivity was rationalized as arising from hydrogen bonding

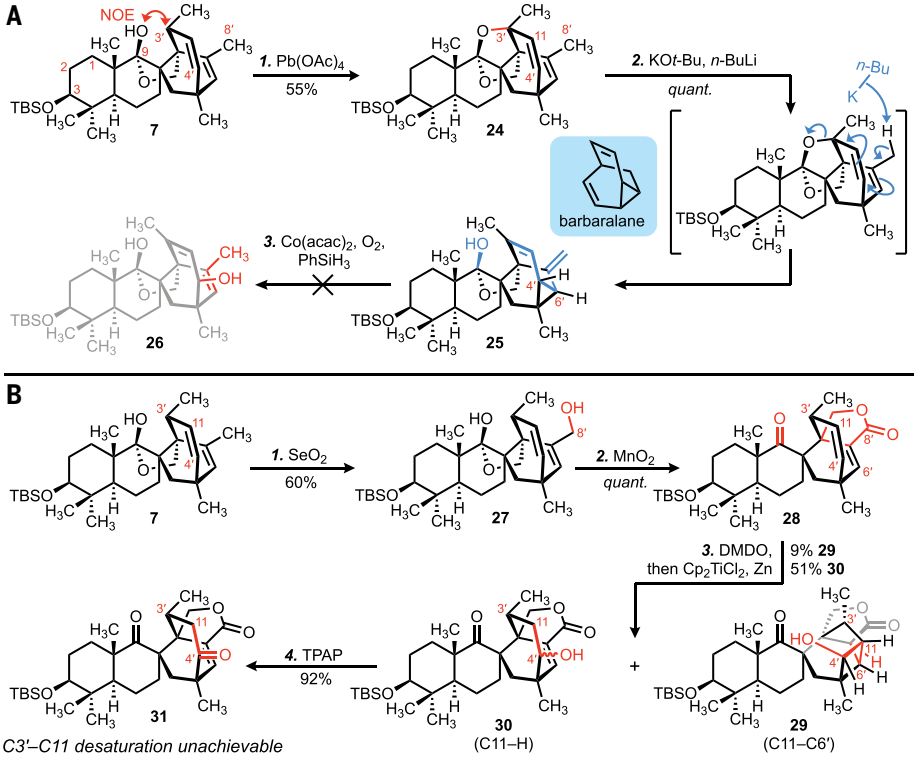


Fig. 3. Attempts to oxidize or desaturate the C3′ position. (**A**) Intramolecular hydrogen abstraction. Reagents and conditions: (1) lead tetraacetate [$Pb(OAc)_4$], benzene, 90°C; (2) potassium *tert*-butoxide (KO*t*-Bu), *n*-butyllithium (*n*-BuLi), THF, −78°C; (3) cobalt(II) acetylacetonate [$Co(acac)_2$], phenylsilane ($PhSiH_3$), O_2 (balloon), RT. (**B**) Saegusa-Ito oxidation. Reagents and conditions: (1) selenium dioxide (SeO_2), 1,4-dioxane–pyridine (2:1), 80°C; (2) manganese dioxide (MnO_2), CH_2Cl_2, RT; (3) DMDO, CH_2Cl_2, 0°C, then titanocene dichloride (Cp_2TiCl_2), Zn^0, triethylsilane (Et_3SiH), THF, −78°C to RT; and (4) tetrapropylammonium perruthenate (TPAP), CH_2Cl_2, 0°C.

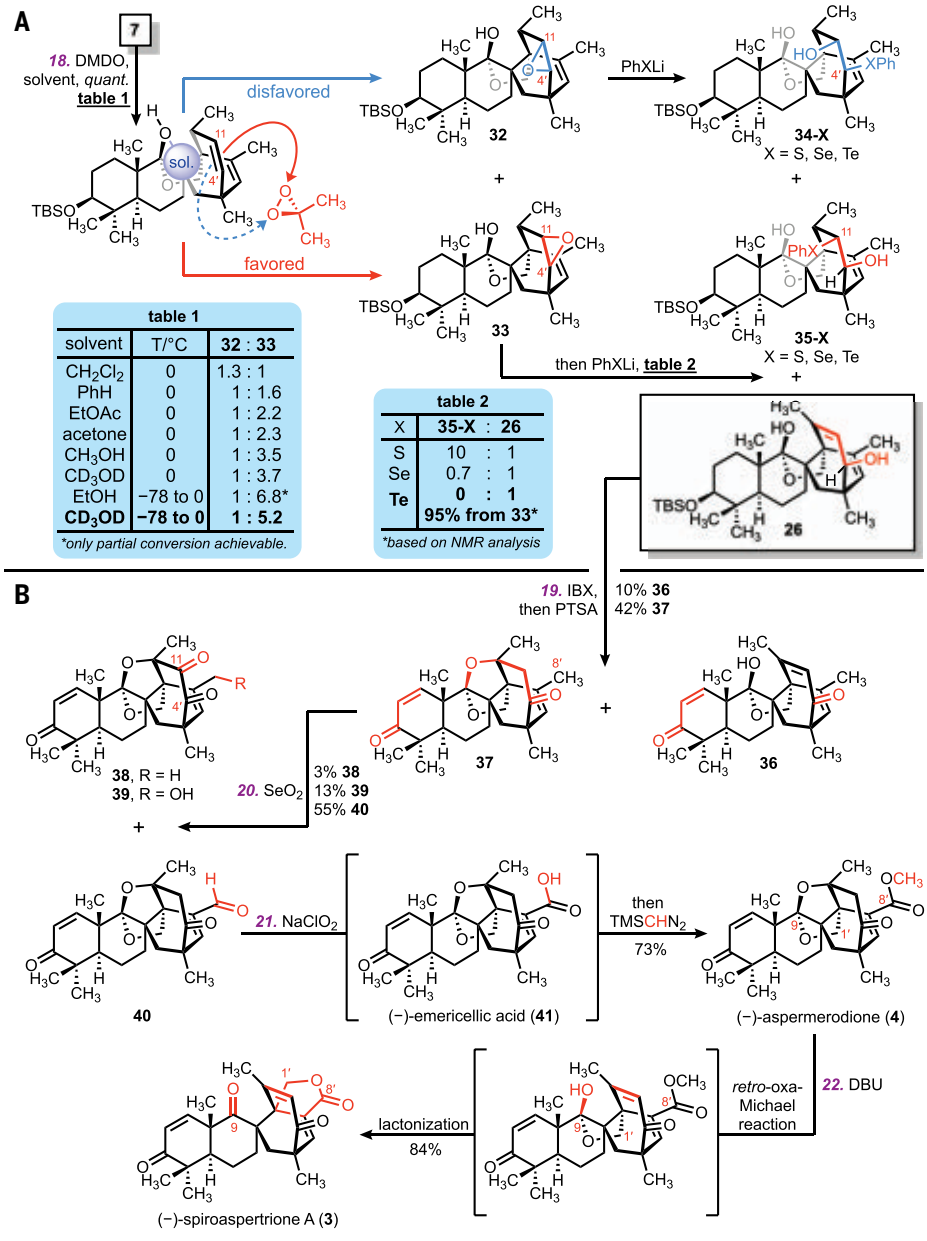


Fig. 4. Late-stage functionalizations. (**A**) Preparation of allylic alcohol **26** through epoxide **30**. Reagents and conditions (step count continued from Fig. 2): (18) DMDO, then lithium phenyltelluride (PhTeLi), 90°C. (**B**) Synthesis of (−)-spiroaspertrione A (**3**). Reagents and conditions: (19) 2-IBX, DMSO, RT, then PTSA, 70°C; (20) SeO_2, 1,4-dioxane, 120°C; (21) sodium chlorite ($NaClO_2$), monosodium phosphate (NaH_2PO_4), 2-methyl-2-butene, *t*-BuOH–THF–H_2O (1:2:1), 0°C, then trimethylsilyldiazomethane ($TMSCHN_2$); and (22) DBU, benzene, 100°C.

between the C9 hemiketal and the protic solvent, which serves to occlude the approach of DMDO to the undesired face of the C11–C4′ alkene. Therefore, using appropriately sized alcohols and lower temperatures should in principle favor the formation of epoxide ***33***. Indeed, deuterated methanol slightly improved the results, whereas lowering the temperature significantly boosted the ratio of ***33*** to ***32***. Even better diastereoselectivity was observed in ethanol, but its vulnerability to DMDO oxidation made full conversion challenging to achieve. To increase the amount of **26**, we varied the base (Fig. 4A and table S2). Heating ***33*** with lithium phenyltelluride (PhTeLi) at 90°C generated **26** as the sole product in almost quantitative yield, whereas lithium phenylthiolate (PhSLi) predominantly produced **35-S**.

The allylic alcohol **26** was oxidized with IBX in dimethyl sulfoxide (DMSO) in the presence of para-toluenesulfonic acid (PTSA) (*28*) to provide enone **36** and the bicyclic ketal **37**. Allylic oxidation of the C8′ methyl group of **37** was performed with selenium dioxide, yielding aldehyde **40**, along with **38** and **39**, arising from Riley oxidation (*29*) of the ketone C4′. The pronounced tendency for enolization at the C11 position, however, supported our earlier hypothesis regarding the cascade transformation from (−)-aspermerodione (**4**) to (−)-spiroaspertrione A (**3**). The preparation of **4** was achieved from aldehyde **40** through Pinnick oxidation (*30*) and trapping of the in situ generated (−)-emericellic acid (**41**) using $TMSCHN_2$. As hypothesized, heating with 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) at 100°C triggered a *retro*-oxa-Michael reaction, leading to the collapse of the C9 hemiketal and liberating the C1′ primary alcohol, which subsequently attacked the C8′ methyl ester, thus forming the lactone moiety of (−)-spiroaspertrione A (**3**). The optical rotations of both synthetic

natural products ***3*** and ***4*** matched the values reported in the literature (*6*, *7*).

Conclusions

We accomplished the enantioselective total synthesis of an andiconin-derived meroterpenoid, (−)-spiroaspertrione A (***3***), in 16 steps from the literature-reported compound **11** with an overall yield of 2.3%. This achievement allowed us to study the previously hypothesized interconversion between (−)-spiroaspertrione A (***3***) and (−)-aspermerodione (***4***). The synthesis also granted us broader insight into the relationship between the more common frameworks of the andiconin family.

REFERENCES AND NOTES

1. A. W. Dunn, R. A. W. Johnstone, B. Sklarz, T. J. King, *J. Chem. Soc. Chem. Commun.* **270**, 270a (1976).
2. R. Geris, T. J. Simpson, *Nat. Prod. Rep.* **26**, 1063–1094 (2009).
3. Y. Matsuda, I. Abe, *Nat. Prod. Rep.* **33**, 26–53 (2016).
4. M. Jiang, Z. Wu, L. Liu, S. Chen, *Org. Biomol. Chem.* **19**, 1644–1704 (2021).
5. Y. Matsuda, T. Wakimoto, T. Mori, T. Awakawa, I. Abe, *J. Am. Chem. Soc.* **136**, 15326–15336 (2014).
6. Y. He *et al.*, *J. Org. Chem.* **82**, 3125–3131 (2017).
7. Y. Qiao *et al.*, *Sci. Rep.* **8**, 5454 (2018).
8. T. Bai *et al.*, *Org. Lett.* **22**, 4311–4315 (2020).
9. E. Vogel, *Angew. Chem.* **72**, 4–26 (1960).
10. E. Vogel, K. H. Ott, K. Gajek, *Justus Liebigs Ann. Chem.* **644**, 172–188 (1961).
11. J. R. Falck, S. Manna, S. Chandrasekhar, L. Alcaraz, C. Mioskowski, *Tetrahedron Lett.* **35**, 2013–2016 (1994).
12. M. E. Jung, D. Ho, H. V. Chu, *Org. Lett.* **7**, 1649–1651 (2005).
13. I. Ryu, S. Murai, Y. Hatayama, N. Sonoda, *Tetrahedron Lett.* **19**, 3455–3458 (1978).
14. Y. Morita, M. Suzuki, R. Noyori, *J. Org. Chem.* **54**, 1785–1787 (1989).
15. E. L. Fisher, S. M. Wilkerson-Hill, R. Sarpong, *J. Am. Chem. Soc.* **134**, 9946–9949 (2012).
16. H. M. Davies, E. Saikali, W. B. Young, *J. Org. Chem.* **56**, 5696–5700 (1991).
17. W. von E. Doering, A. K. Hoffmann, *J. Am. Chem. Soc.* **76**, 6162–6165 (1954).
18. M. Mąkosza, M. Wawrzyniewicz, *Tetrahedron Lett.* **10**, 4659–4662 (1969).
19. J. B. Baudin, G. Hareau, S. A. Julia, O. Ruel, *Tetrahedron Lett.* **32**, 1175–1178 (1991).
20. K. Heusler, J. Kalvoda, *Angew. Chem. Int. Ed.* **3**, 525–538 (1964).
21. M. L. Mihailović, Ž. Čeković, *Synthesis (Stuttg.)* **1970**, 209–224 (1970).
22. K. Fujita, M. Schlosser, *Helv. Chim. Acta* **65**, 1258–1263 (1982).
23. M. Schlosser, *Pure Appl. Chem.* **60**, 1627–1634 (1988).
24. A. Zech, C. Jandl, T. Bach, *Angew. Chem. Int. Ed.* **58**, 14629–14632 (2019).
25. S. Isayama, T. Mukaiyama, *Chem. Lett.* **18**, 1071–1074 (1989).
26. Y. Ito, T. Hirao, T. Saegusa, *J. Org. Chem.* **43**, 1011–1013 (1978).
27. S. V. Ley, J. Norman, W. P. Griffith, S. P. Marsden, *Synthesis (Stuttg.)* **1994**, 639–666 (1994).
28. K. C. Nicolaou, T. Montagnon, P. S. Baran, *Angew. Chem. Int. Ed.* **41**, 993–996 (2002).
29. H. L. Riley, J. F. Morley, N. A. C. Friend, *J. Chem. Soc.* 1875–1883 (1932).
30. B. S. Bal, W. E. Childers Jr., H. W. Pinnick, *Tetrahedron* **37**, 2091–2096 (1981).
31. W. Huang, L. Pan, H. Zhao, F. Schneider, T. Gaich, Data for: The total synthesis of (−)-spiroaspertrione A: A divinylcyclopropane rearrangement approach, Zenodo (2025); https://doi.org/10.5281/zenodo.16889403.

ACKNOWLEDGMENTS

We thank J. R. Falck (UT Southwestern, Dallas), S. Chandrasekhar (GITAM, Hyderabad), and S. Kallepu (Moffitt Cancer Center) for granting us their original experimental procedure for the flash vacuum pyrolysis; S. B. Herzon (Yale University) and D. Sarlah (Rice University) and H. Zhao and S. Ma (University of Konstanz) for helpful discussions during the project; M. Kovermann, A. Friemel, and U. Haunz (NMR department of the University of Konstanz) for assistance with NMR spectroscopy; M. Bein (University of Konstanz) for assistance with high-resolution mass spectroscopy; and the staff of the SCCKN (Scientific Computer Cluster in Konstanz) at the University of Konstanz for HPC resources. **Funding:** This work was supported by the European Research Council (ERC-consolidator grant ERC-COG-CoSyMoDe, "Cooperative Synthesis by Molecular Deconvolution," grant no. 101003072 to T.G.) and Shanghai University of Traditional Chinese Medicine (Shanghai Bay talent program and research start-up fund for introduced talents to F.S.). **Author contributions:** Conceptualization: T.G., F.S.; Data collection: W.H.; DFT calculations: H.Z.; Funding acquisition: T.G.; Investigation: W.H., F.S.; Writing – original draft: F.S.; Writing – review & editing: F.S., T.G., W.H., L.P.; Writing – supporting information: W.H. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. Free induction decay values of NMR spectra of compounds **3**, **4**, **7**, **(*E*)-8**, **17**, **22**, **24**, **25**, **26**, **28**, **37**, and **40** are accessible at Zenodo (*31*).

License information: Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adz7593
Materials and Methods; Spectral Data; References (*32*–*42*)

Submitted 13 June 2025; accepted 21 August 2025

10.1126/science.adz7593

INFECTIOUS DISEASE

Architectural immunity: Ants alter their nest networks to prevent epidemics

Luke Leckie[1,2,3]*, Mischa Sinha Andon[1], Katherine Bruce[1], Nathalie Stroeymeyt[1]*

In animal groups, spatial structure shapes social interaction patterns, thereby influencing the transmission of infectious diseases. Active modifications to the spatial environment could therefore be a potent tool to mitigate epidemic risk. We tested whether *Lasius niger* ants modify their nest architecture in response to pathogens by introducing control- or pathogen-treated individuals into nest-digging groups and monitoring three-dimensional nest morphogenesis. Pathogen exposure led to architectural changes, including faster nest growth, increased interentrance distance, transmission-inhibitory changes in nest network topology, and reduced chamber centrality. Simulations confirmed that these changes reduced transmission and highlighted a synergy between architectural and behavioral responses to disease. These results provide evidence for architectural immunity in a social animal and offer insights into how spatial organization can be leveraged to decrease epidemic susceptibility.

Animal and human networks influence the spread of disease from local to global scales. For instance, the properties of contact and social networks determine the risk and severity of epidemics within social groups (*1*–*3*); urban and metapopulation networks shaped by the layout of buildings, cities, and natural habitats influence social interaction and dispersal patterns and thus affect disease transmission dynamics (*4*–*7*); and global transportation networks facilitate the long-range transmission of disease by connecting distant populations (*8*).

There is increasing evidence that modifying social contact networks is an effective intervention strategy against epidemics that is used in a broad range of animal and human societies. Several species, including ants, humans, guppies, mice, and mandrills, are known to avoid infected conspecifics, thereby reducing disease transmission rates (*9*–*11*). Furthermore, ants and humans respond to pathogens by increasing the compartmentalization of their social networks to limit pathogen spread across the group (*2*, *12*). In addition to modifying their social networks, human societies have also used modifications to their spatial networks as an active means to reduce disease transmission (*13*). Historical examples include the expansion of urban spaces and the separation of cities into functional zones as preventative measures against outbreaks of the bubonic plague in the 1300s and cholera in the 1800s, respectively (*14*, *15*). More recently, active network–based interventions have been proposed that target the spatial properties of transport networks, city layouts, and building architecture to control modern human pandemics, such as COVID-19 (*6*, *14*, *16*, *17*). Yet, because of the lack of empirical data, it is still unclear which spatial interventions may be most effective at limiting pathogen spread while preserving the functioning of society. Animal societies that inhabit complex built structures could provide an opportunity to study evolved solutions to this challenge. However, there is, as yet, no evidence that nonhuman animals actively modify their spatial surroundings to mitigate epidemic risk, even though such interventions could be highly effective.

In this work, we investigate the possible role of architectural changes in nest layout as a disease defense strategy in the black garden ant *Lasius niger*. Nest-building social insects are an ideal system to explore for evolved architectural, disease-targeting interventions in animal societies. Because high interaction rates between related colony members favor the transmission of infectious pathogens, social insects have evolved a large suite of collective mechanisms for disease defense that confer "social immunity," including active modifications to colony social interaction networks (*2*, *18*). Furthermore, excavated ant nests can demonstrate a high degree of complexity, with specialized chambers housing food, brood, reproductive individuals, or waste connected by tunnels into underground networks that successfully isolate potential infectious sources (*19*, *20*). However, although previous work has indicated that excavation dynamics are influenced by abiotic stressors, such as temperature and soil composition (*21*–*25*), and by the presence of fungal spores in the soil (*26*), it is currently unknown whether ants respond to infectious threats by actively modifying their whole-nest architecture to reduce disease transmission. In this work, we test this hypothesis using a combination of pathogen exposure, video recording, x-ray micro–computed tomography (micro-CT), spatial network analysis, and agent-based simulations of disease transmission within nests.

Results

To investigate how pathogen exposure influences nest digging by *L. niger* ants, we allowed groups of 180 workers to excavate a new nest in a digging arena. One day after the onset of excavation, we introduced 20 additional workers, either exposed to the generalist entomopathogenic fungus *Metarhizium brunneum* (pathogen-exposed nests, $n = 10$) or treated with a sham solution (control nests, $n = 10$), into the digging arena. *M. brunneum* is a natural ant pathogen (*27*, *28*) that is picked up from the environment by contact with sporulating cadavers. Contaminated workers can transmit infectious spores to nestmates through physical contact for up to 1 day after exposure, ~90% of directly exposed workers die from infection within 10 days, and the mortality of nestmates increases by a factor of ~3 (*2*). After the introduction of the treated ants, we monitored nest excavation for 6 days using video recording of surface activity and nondestructive micro-CT scans of the internal nest structure (Fig. 1, A and B; movie S1; and materials and methods).

Pathogen exposure influences surface activity and surface properties of the nest

To investigate the influence of treatment on individual-level activity at the surface, we recorded the number of treated and untreated workers leaving the nest via the main (central) entrance (materials and methods). The exit rate of both untreated and treated workers decreased significantly over time in both treatments. However, whereas the exit rate of untreated workers was unaffected by treatment, treated workers exited the main nest entrance at a significantly higher rate in pathogen-exposed compared with control nests [fig. S1 and table S1; linear mixed-effects model (LMM), effect of time: untreated workers, $\chi^2 = 208.74$, df = 1, $P < 0.0001$; treated workers, $\chi^2 = 6.30$, df = 1, $P = 0.012$; effect of treatment: untreated workers, $\chi^2 = 0.021$, df = 1, $P = 0.89$; treated workers, $\chi^2 = 6.53$, df = 1, $P = 0.011$; interaction treatment × time: both untreated and treated workers, $\chi^2 \leq 2.11$, $P \geq 0.14$]. This indicates that pathogen exposure increases the surface activity of directly treated workers but not their nestmates. Because pathogen exposure does not lead to increased activity in this host-pathogen system (fig. S2), this likely reflects self-isolation by exposed workers outside the nest, as has already been reported in several species of ant, including *L. niger* (*2*, *11*, *29*). Furthermore, we found that entrances were spaced further apart from one another in pathogen-treated versus control nests (Fig. 2; LMM, effect of treatment throughout

[1]School of Biological Sciences, University of Bristol, Bristol, UK. [2]Department of Informatics, Luddy School of Informatics, Computing, and Engineering, Indiana University Bloomington, Bloomington, IN, USA. [3]Cognitive Science Program, Indiana University Bloomington, Bloomington, IN, USA. *Corresponding author. Email: lleckie@iu.edu (L.L.); nathalie.stroeymeyt@bristol.ac.uk (N.S.)

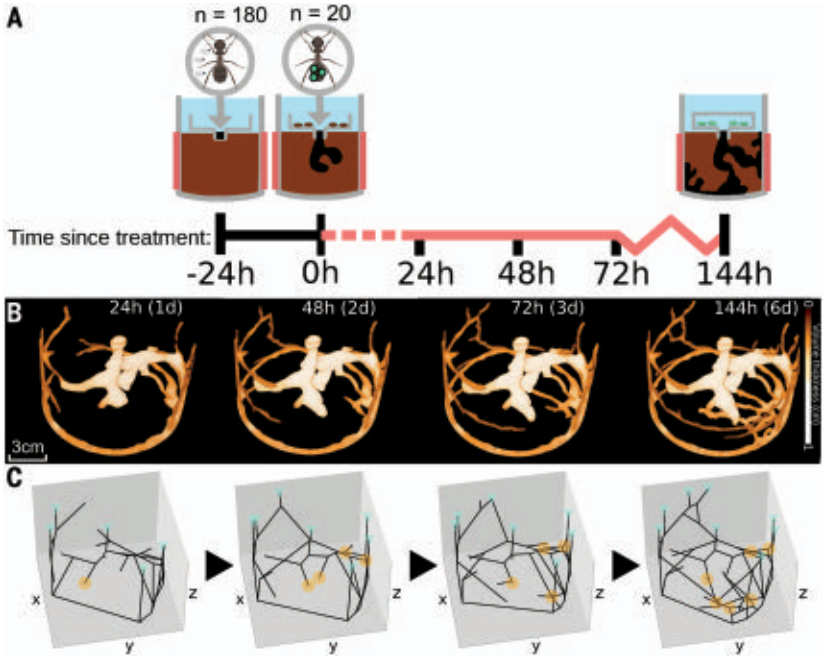


Fig. 1. Experimental protocol and nest network extraction. (**A**) One hundred and eighty ants with 25 mg of early-instar brood and 15 pupae were introduced into a soil-filled container (digging arena) and allowed to excavate a nest freely (−24 hours). After 24 hours, a CT scan of the nest was taken (baseline scan), and then 20 sham- or 20 pathogen-treated ants were introduced near the nest entrance (0 hour). Colonies were allowed to excavate the nest for 6 additional days, and CT scans were taken 24, 48, 72, and 144 hours (1, 2, 3, and 6 days) after treatment. Previous research has revealed no effects of repeated CT scanning on the excavation behavior of *Lasius* ants in a similar experimental setup (*22*). (**B**) CT scan reconstructions of nest volumes 1 day (24 hours), 2 days (48 hours), 3 days (72 hours), and 6 days (144 hours) after treatment. Volume thickness is encoded by increasing color brightness. (**C**) Three-dimensional (3D) spatial networks automatically extracted from the nest CT scans. Cyan triangles indicate nest entrances, orange circles indicate nest chambers, and black lines indicate tunnels (network edges).

experiment, $\chi^2 = 4.50$, df = 1, $P = 0.034$; Fig. 3A; LMM, effect at 6 days: entrances 0.62 ± 0.30 cm further apart in pathogen-treated nests, $\chi^2 = 5.93$, df = 1, $P = 0.015$). Increased spacing between entrances could lead to decreased contact rates between individuals at the surface and has been used as a strategy to mitigate disease transmission in human buildings (*30*).

Pathogen exposure influences overall nest growth

To study the influence of pathogen exposure on nest growth, we automatically identified nest chambers, entrances, tunnels, junctions, and dead ends from the micro-CT scans and measured their volumes (Fig. 1 and materials and methods). Analyses of nest properties in pre-exposure scans (0 hour) (i.e., before the introduction of treated ants) revealed no significant between-treatment differences, which indicates that any differences in post-exposure nests should be a result of pathogen exposure rather than random initial fluctuations between treatment groups (table S2).

Pathogen-exposed *L. niger* workers typically sharply decrease the amount of time that they spend inside the nest after exposure and slightly reduce their activity levels (fig. S2). Hence, if nest architecture changes as a passive consequence of infection, we would expect pathogen-exposed nests to grow more slowly than control nests (*31*). Conversely, because higher nest volumes should lead to lower ant density and thus fewer physical contacts, if the ants actively adjust their nest architecture to decrease transmission risk, we would expect pathogen-exposed groups to excavate nests at a faster rate. In agreement with the latter prediction, we found that the overall nest volume increased significantly faster in pathogen-exposed versus control nests (fig. S3; LMM, interaction treatment × time, $\chi^2 = 5.78$, df = 1,

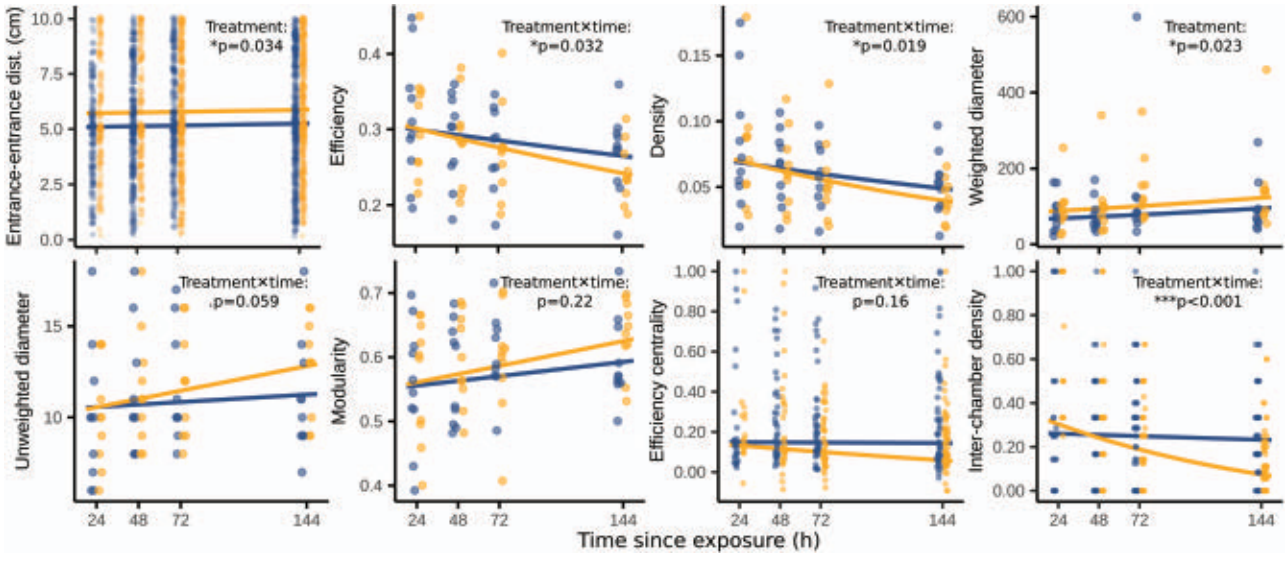


Fig. 2. Nest and chamber architectural properties as a function of time. Lines represent LMM fits for control (blue) and pathogen-treated colonies (orange), back-transformed where appropriate. Significance of the main treatment effect or the interaction between treatment and time (treatment × time) on a nest property are indicated on each graph. Original *P* values are provided, with an estimated study-wide false discovery rate (expected proportion of false positive results across all separate analyses) of 0.072 (~1 out of 14). For entrance-entrance distance, each point represents a pairwise measure of Euclidean distance between a new entrance and any other entrance ($n = 2529$ pairs). For efficiency, density, weighted diameter, unweighted diameter, and modularity, each point represents one nest ($n = 79$). For efficiency centrality and interchamber density, each point represents one chamber ($n = 336$). Efficiency, density, weighted diameter, efficiency centrality, and interchamber density were log-transformed for statistical analyses. Definitions of all properties are provided in Table 1, and model coefficients and exact *P* values are provided in table S1.

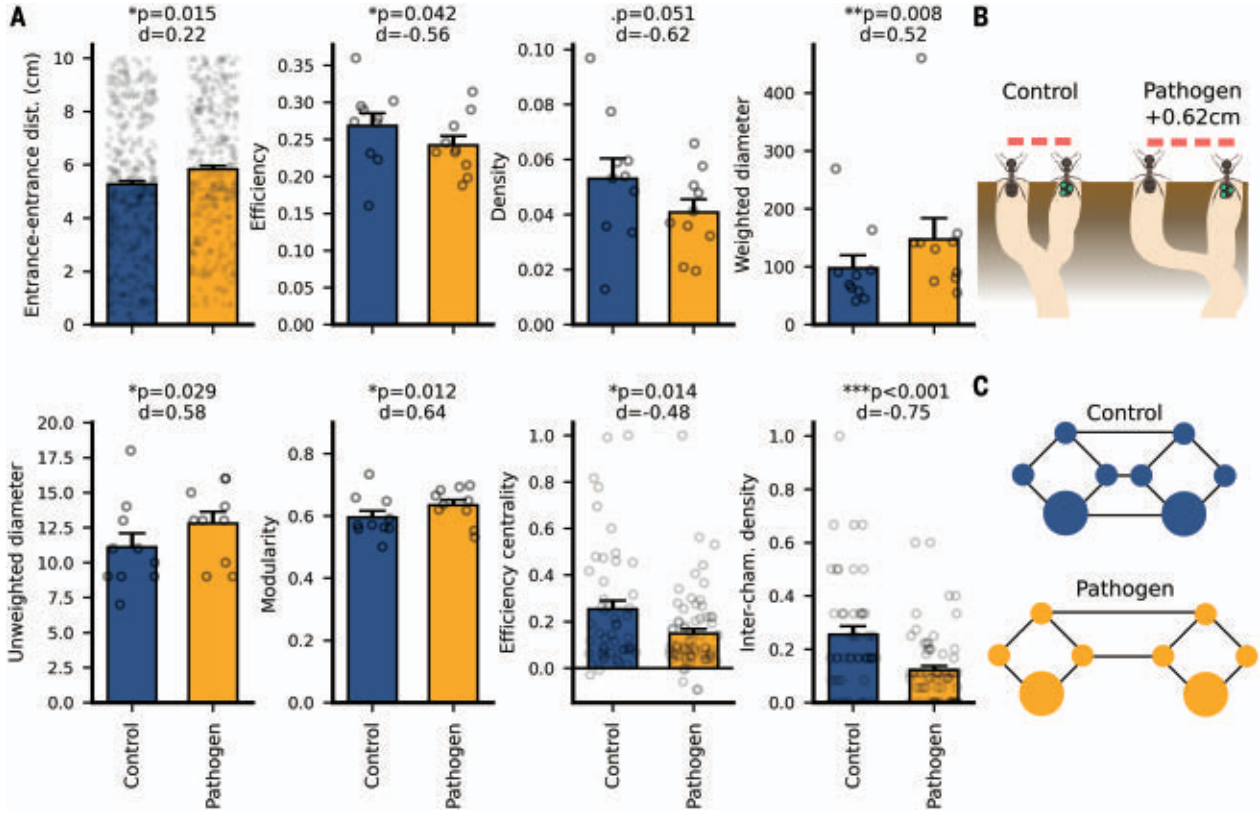


Fig. 3. Pathogen-induced architectural modifications 6 days after treatment. (**A**) Bars and whiskers show the means and standard errors for each property. Points represent individual data points [entrance-entrance distance (dist.): $n = 2529$ pairs of entrances; efficiency, density, weighted diameter, unweighted diameter, and modularity: $n = 20$ nests; efficiency centrality, interchamber density: $n = 336$ chambers]. Significance of the main effect of treatment at day 6 (LMM) and Cohen's d effect size (|d| ~ 0.2: weak effect; |d| ~ 0.5: medium effect; |d| ~ 0.8: strong effect) are indicated above each plot. Model coefficients and exact *P* values are provided in table S1. Original *P* values are provided, with an estimated study-wide false discovery rate (expected proportion of false positive results across all separate analyses) of 0.072 (~1 out of 14). (**B**) Illustration of entrance spacing mediating reduced interaction rate (not to scale). Entrances in pathogen-exposed nests (right) were spaced on average 0.62 cm (~1 to 1.5 *L. niger* worker body lengths) further apart compared with control nests (left), which could reduce the rate of interactions at the surface. (**C**) Simplified illustration of between-treatment differences in control (blue) and pathogen-exposed (orange) architectures after 6 days. Small circles represent junctions, large circles represent chambers, and black lines represent tunnels. Pathogen-exposed nests had higher (weighted and unweighted) diameter and modularity but lower density and network efficiency, and chambers had lower efficiency centrality and interchamber density.

$P = 0.016$), although this effect was too small to translate into significant between-treatment differences in nest volume by the end of the experiment (effect of treatment at 6 days, $\chi^2 = 0.23$, df = 1, $P = 0.63$). The faster growth of pathogen-exposed nests was not a result of a faster increase in the number and/or volume of nest chambers or in the number of nest entrances (fig. S3; LMM, interaction treatment × time, $\chi^2 \leq 1.80$, df = 1, $P \geq 0.18$, for all variables). Instead, pathogen exposure led to a significant increase in the rate of tunnel formation (fig. S3; LMM, interaction treatment × time, $\chi^2 = 5.49$, df = 1, $P = 0.019$), which suggests that it may affect the connectivity of the nest network.

Pathogen exposure influences the overall nest network topology

To test whether pathogen-exposed groups of *L. niger* workers alter the topology of their nest network to decrease epidemic risk, we identified properties known to influence disease transmission in social networks (*1*, *2*, *32*) (Table 1) and measured them in the spatial nest networks consisting of chambers, junctions, dead ends, and entrances (nodes) connected by tunnels (edges; Fig. 1C). We predicted that pathogen-exposed nest networks would display increased transmission-inhibitory properties and decreased transmission-enhancing properties compared with control nests.

In agreement with this prediction, we found that nest network efficiency and density (transmission-enhancing properties) decreased at a significantly higher rate in pathogen-exposed compared with control nests (Fig. 2; LMM, effect of treatment × time: efficiency, $\chi^2 = 4.58$, df = 1, $P = 0.032$; density, $\chi^2 = 5.52$, df = 1, $P = 0.019$). Both properties were reduced in pathogen-exposed nests by 6 days after treatment (Fig. 3A; LMM, effect of treatment at 6 days: efficiency, $\chi^2 = 4.125$, df = 1, $P = 0.042$; density, $\chi^2 = 3.82$, df = 1, $P = 0.051$). In addition, we found that the weighted diameter of the nest network (transmission-inhibitory property) was significantly higher in pathogen-exposed compared with control nests throughout the experiment (Fig. 2; LMM, effect of treatment throughout experiment, $\chi^2 = 5.17$, df = 1, $P = 0.023$; Fig. 3A; LMM, effect at 6 days, $\chi^2 = 6.97$, df = 1, $P = 0.008$). Furthermore, there were nonsignificant trends for the unweighted diameter and modularity of nest networks (transmission-inhibitory properties) to increase faster in pathogen-exposed versus control nests (Fig. 2; LMM, effect of treatment × time: unweighted diameter, $\chi^2 = 3.57$, df = 1, $P = 0.059$; modularity, $\chi^2 = 1.48$, df = 1, $P = 0.22$), and both properties were significantly higher in pathogen-exposed nests by the end of the experiment (Fig. 3A; LMM, effect of treatment at 6 days: unweighted diameter, $\chi^2 = 4.79$, df = 1, $P = 0.029$; modularity, $\chi^2 = 6.34$, df = 1, $P = 0.012$). No other transmission-relevant properties of the

Table 1. Summary of network properties measured from nest networks and their predicted effects on disease transmission. A plus sign (+) indicates enhancement, and a minus sign (−) indicates inhibition.

Global property	Definition—global network	Effect on transmission
Efficiency	Ability of the network to convey information quickly across all nodes	+ (*45*)
Density	Proportion of existing connections among all possible connections	+ (*46*)
Weighted diameter	Shortest weighted path length (i.e., sum of edge weights) between the two most distant nodes in the network	− (*47*)
Unweighted diameter	Shortest unweighted path length (i.e., number of steps) between the two most distant nodes in the network	− (*47*)
Modularity	Extent to which a network is divided into separate communities with many within-community and few intercommunity connections	− (*48*)
Clustering	Tendency of neighboring nodes to form fully connected sets	Context-dependent effect (*49*)
Degree heterogeneity	Variation in the number of unweighted connections (edges) across all nodes in a network	+ (*8*)
Node property	**Definition—node level**	**Effect on node susceptibility**
Degree	Number of edges connected to a given node	+ (*49*)
Betweenness centrality (random walk)	Extent to which a node acts as an intermediary for the flow traffic between all other pairs of nodes, when agents travel by random walk	+ (*50*)
Efficiency centrality	Spreading ability of a node, measured as the relative drop in network efficiency caused by its removal	+ (*51*)
Interchamber density	Proportion of other chambers directly connected to a focal chamber	+

overall nest network were affected by treatment (table S1; LMM for clustering and degree heterogeneity: effect of treatment × time, $\chi^2 \leq 0.07$, df = 1, $P \geq 0.51$; effect of treatment at 6 days, $\chi^2 \leq 1.35$, df = 1, $P \geq 0.25$, in all tests). Overall, these results indicate that pathogen exposure leads to multiple changes in the overall topology of the nest networks that should decrease pathogen transmission.

Pathogen exposure influences the topological position of nest chambers

Previous studies of transport and social networks have shown that highly populated nodes and nodes with high network centrality (e.g., degree, betweenness, and efficiency centrality; Table 1) are both at high risk of contamination and are highly influential for the onward spread to the rest of the network (*2*, *33*–*35*). Because nest chambers contain a large portion of the nest population, including valuable and vulnerable colony members such as the queen, young adults, and the brood, we predicted that pathogen-exposed groups would build chambers in positions associated with lower exposure risk and lower spreading ability, characterized by lower network centrality and reduced connections to other chambers.

Most chambers had a degree of three, and neither chamber degree nor betweenness differed between treatments [table S1; Poisson GLMM (degree) and LMM (betweenness), interaction time × treatment and main treatment effects, $\chi^2 \leq 0.03$, df = 1, $P \geq 0.59$, in all tests; Poisson GLMM (degree) and LMM (betweenness) at 6 days, $\chi^2 \leq 0.20$, df = 1, $P \geq 0.66$]. However, in agreement with our prediction, the efficiency centrality of nest chambers tended to decrease faster in pathogen-exposed compared with control nests and was significantly reduced in pathogen-exposed nests by the end of the experiment (Fig. 2; LMM, interaction treatment × time, $\chi^2 = 1.97$, df = 1, $P = 0.16$; Fig. 3A; LMM, effect of treatment at 6 days, $\chi^2 = 6.07$, df = 1, $P = 0.014$). Furthermore, the density of connection of chambers to other chambers decreased significantly faster in pathogen-exposed versus control nests and was significantly reduced in pathogen-exposed nests by 6 days after exposure (Fig. 2; LMM, interaction treatment × time, $\chi^2 = 25.30$, df = 1, $P < 0.0001$; Fig. 3A; LMM, effect of treatment at 6 days, $\chi^2 = 16.02$, df = 1, $P < 0.001$). Altogether, these results indicate that pathogen-exposed ants excavate chambers in positions that should reduce the severity of epidemics.

Pathogen-induced changes in nest architecture decrease disease transmission

Our results so far suggest that pathogen exposure triggers a range of transmission-inhibitory changes in nest architecture by 6 days after initial pathogen exposure (Fig. 3), from the self-isolation of exposed workers to the surface to changes in the geometry and topology of the excavated nests. This could play an important role in preventing epidemics upon later, secondary disease challenges, arising for example from contact with the sporulating bodies of the initially exposed individuals or from new infectious sources in pathogen-rich environments. To formally test the impact of these changes on secondary epidemic risk, we developed an agent-based model that simulated the transmission of an infectious pathogen within the observed nest networks at day 6. This model was inspired from previous models of disease transmission in ant nests (*1*, *2*, *32*) and ant traffic in confined space (*36*, *37*) (materials and methods). In the model, agents moved through the nest and aggregated locally inside nest chambers (*20*, *38*). Pathogen-treated agents initially entered the nest through a randomly selected entrance, and pathogen transmission occurred as a stochastic process between pairs of agents sharing the same within-nest location (chambers, tunnels, dead ends, and junctions). We simulated the transmission of *M. brunneum* over the experimental nest networks extracted 6 days after treatment after a similar disease challenge (20 out of 200 agents), as in our experiments. Relative to simulations of transmission through control nests, simulations of transmission through pathogen-exposed nests resulted in a significant reduction in the amount of spores received by untreated agents and in the proportion of these agents that received a high (more lethal) load (Fig. 4A; LMM interaction treatment × time: median load, $\chi^2 = 102.546$, df = 1, $P < 0.0001$; prevalence of high load, $\chi^2 = 127.24$, df = 1, $P < 0.0001$). Furthermore, we found that the architecture of pathogen-exposed nests inhibited transmission even when only a single pathogen-treated agent was introduced (fig. S4A and table S3), demonstrating the robustness of this effect.

To tease apart the relative importance of different pathogen-induced architectural modifications in reducing transmission, we applied partial least-squares regression to the outcome of simulations over all nests. All properties were found to have the expected effects on transmission: Presumed transmission-inhibitory properties (weighted diameter, unweighted diameter, modularity, and interchamber density)

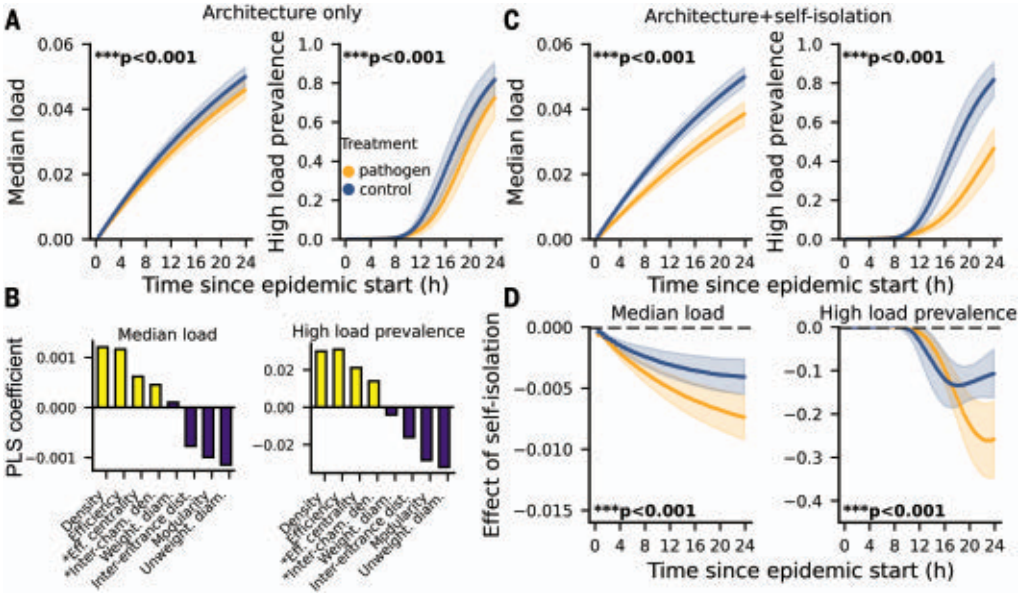


Fig. 4. Agent-based simulations of *M. brunneum* transmission within 6-day nest networks. (**A** and **B**) Simulation outcomes without self-isolation in either treatment group. (**C**) Simulation outcomes with self-isolation of treated agents in pathogen-exposed nests only. (**D**) Difference in simulated transmission with self-isolation versus without self-isolation in both treatment groups. One hundred simulations were performed on each nest for each condition (with or without self-isolation), and cross-simulation means were calculated for each nest every 800 s ($n = 107$ time points), resulting in one data point per nest per time point and per condition. [(A) and (C)] Points and shaded areas represent the grand means and standard errors of the median load (left) and high load prevalence (right) across untreated agents in simulations over control (blue, $n = 10$ per time point) and pathogen-exposed (orange, $n = 10$ per time point) nest networks. (B) Bars indicate coefficients from partial least-squares regression (PLS) analyzing the relative effects of nest network properties targeted by architectural modifications on simulated median load (left) and high load prevalence (right) in the absence of any self-isolation. Positive (resp. negative) values indicate a positive (resp. negative) association between a property and transmission, and bar height is proportional to its relative influence on transmission. Transmission-enhancing properties are indicated in yellow, and transmission-inhibitory properties are indicated in purple. Chamber properties are highlighted by an asterisk; all other properties apply to the global topology of the nest. All pathogen-induced changes were transmission inhibitory. (D) Points and shaded areas represent the grand means and standard errors of the cross-condition differences in median load and high load prevalence (transmission with self-isolation minus transmission without self-isolation; $n = 10$ nests per treatment and per time point). Negative values reflect a transmission-inhibitory effect of self-isolation by treated agents. Details of statistical analyses are provided in table S3.

were negatively associated with transmission, whereas presumed transmission-enhancing properties (density, efficiency, and efficiency centrality) were positively associated with transmission (Fig. 4B). Efficiency, density, unweighted diameter, and modularity (whole-network properties) had the strongest associations with transmission; chamber properties had weaker associations, with efficiency centrality showing a stronger effect on transmission compared with interchamber density; and weighted nest diameter and interentrance distances had the weakest association with transmission (Fig. 4B).

Previous research (*2*) and our surface-activity results (fig. S1) indicate that infectious ants isolate themselves from unexposed nestmates, which reduces transmission risk. To investigate potential interactions between this behavioral response and architectural responses to disease, we repeated our simulations with an additional self-isolation mechanism, whereby pathogen-treated agents displayed an upward movement bias (i.e., increased probability of selecting tunnels connected to higher elevation nodes, such as nest entrances) that led them to spend more time outside the nest (figs. S5 to S7 and table S3). We found that the combined effects of pathogen-induced architectural changes and self-isolation in pathogen-exposed nests resulted in much stronger inhibition of disease transmission compared with architecture alone (Fig. 4C; LMM interaction treatment × time: median load, $\chi^2 = 279.00$, df = 1, $P < 0.0001$; prevalence of high load, $\chi^2 = 212.241$, df = 1, $P < 0.0001$). This was not due to purely additive effects between the two types of responses. Instead, self-isolation was significantly more effective at inhibiting transmission in pathogen-exposed versus control nest architectures (Fig. 4D; LMM interaction treatment × self-isolation × time: median load, $\chi^2 = 37.77$, df = 1, $P < 0.0001$; prevalence of high load, $\chi^2 = 15.19$, df = 1, $P < 0.001$), demonstrating a positive synergy between architectural and behavioral responses, which was robust across a broad range of simulation parameters (fig. S8). Further analyses indicated that this synergy may be a consequence of pathogen-exposed nest networks facilitating the self-isolation of infectious individuals because the same individual-level movement rules (local upward movement bias) tended to result in greater spatial (proportion of time spent outside) and social (network distance between healthy and treated agents) distancing in pathogen-exposed nests (fig. S5 and table S3). Altogether, these results suggest that although the direct effects of pathogen-induced architectural changes on disease transmission are relatively small (Fig. 4A), their indirect effects through synergy with self-isolation are much greater (Fig. 4, C and D), providing enhanced protection against pathogen threats.

Discussion

Group-living animals have evolved collective responses to infectious pathogens that decrease the risk of epidemics (*2*, *9*–*12*). Most responses reported so far involve changes in social interactions between individuals, ranging from social care to self-isolation, social distancing, and social network reorganization (*38*). Our results suggest that these social changes are complemented by spatial modifications to the environment that confer protection against future disease challenges, which had previously only been shown in humans.

Most pathogen-induced changes in architecture increased over time, which could be linked with the nest's development. At the beginning of the experiment, nests were small, which may have limited the scope for implementing architectural changes. At this early stage, self-isolation by pathogen-exposed individuals may therefore be the most effective strategy to prevent the spread of infectious material into the nest. As nests grow larger, with more chambers, junctions, and tunnels, they may become more amenable to topological manipulation, making architectural modification a more potent—and durable—epidemic defense strategy. In fact, the number of possible network configurations scales nonlinearly with the number of nodes (*39*, *40*). By emphasizing social distancing early on and architectural modifications later in nest development, pathogen-exposed ants may dynamically adapt their social immunity to implement the most effective defense. Furthermore, by adopting a transmission-inhibitory nest architecture only after experiencing a real disease challenge, colonies may ensure that efficient flows of resources and information are maintained

throughout the nest in disease-free environments (*41*, *42*). These hypotheses highlight the need to examine pathogen-induced architectural modifications in more natural settings because the limited space and time available in our experimental setup may have curtailed the ability of pathogen-exposed colonies to fully express far-reaching transmission-inhibitory nest modifications.

The mechanisms driving the pathogen-induced changes in nest architecture remain to be established. Although the architectural modifications reported in this work could at least in part arise as an incidental side effect of infection symptoms in diseased workers, it is unlikely to be the sole mechanism at play for three reasons. First, changes to entrance-entrance distance and weighted diameter were detectable as early as 24 hours after exposure. Because *M. brunneum* takes more than 24 hours to enter the host's body and start replicating (*43*), these early changes occurred before the beginning of infection in pathogen-exposed workers and therefore likely reflect an active host response. Second, nests were excavated faster in pathogen-exposed compared with control groups, despite pathogen-exposed workers typically spending less time inside the nest and tending to display reduced activity after exposure, which should decrease excavation rates. This suggests that other mechanisms may be involved. Third, pathogen-treated workers spend much more time near the surface and should thus have less influence on belowground nest morphology compared with other workers. Thus, the topological changes may well involve active shifts in decision-making among untreated workers about where to excavate entrances, chambers, and tunnels. Identifying the fine-grained, individual-level mechanisms that lead to this shift in collective decision-making represents an exciting area of future research.

Overall, our results suggest that ants adjust their nest architecture to reduce colony-wide disease transmission, providing a form of "architectural immunity" that could act as a key component of ant social immunity in pathogen-rich environments (*18*). In addition to uncovering a previously unknown cooperative strategy for avoiding epidemics, our findings provide an example of how social animals may leverage the structure of their built environment to improve their resilience against extrinsic stress, such as extreme temperatures (*24*). Our findings could also have implications beyond animal societies—the architectural changes highlighted in this work have been tuned for effectiveness over long evolutionary time and could serve as a proof of concept or source of inspiration for real-world disease control interventions (*2*, *13*, *44*).

REFERENCES AND NOTES

1. L. Danon, T. A. House, J. M. Read, M. J. Keeling, *J. R. Soc. Interface* **9**, 2826–2833 (2012).
2. N. Stroeymeyt *et al.*, *Science* **362**, 941–945 (2018).
3. I. de Freslon, B. Martínez-López, J. Belkhiria, A. Strappini, G. Monti, *Appl. Anim. Behav. Sci.* **213**, 47–54 (2019).
4. J. A. Tracey, S. N. Bevins, S. VandeWoude, K. R. Crooks, *Ecosphere* **5**, 119 (2014).
5. G. C. Silva, E. M. S. Ribeiro, *Sci. Rep.* **13**, 2240 (2023).
6. S. Uddin, A. Khan, H. Lu, F. Zhou, S. Karim, *Int. J. Environ. Res. Public Health* **19**, 2039 (2022).
7. M. J. Keeling, L. Danon, M. C. Vernon, T. A. House, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 8866–8870 (2010).
8. V. Colizza, A. Barrat, M. Barthélemy, A. Vespignani, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 2015–2020 (2006).
9. J. F. Stephenson, S. E. Perkins, J. Cable, *J. Anim. Ecol.* **87**, 1525–1533 (2018).
10. P. C. Lopes, P. Block, B. König, *Sci. Rep.* **6**, 31790 (2016).
11. N. Bos, T. Lefèvre, A. B. Jensen, P. d'Ettorre, *J. Evol. Biol.* **25**, 342–351 (2012).
12. G. Heiler *et al.*, in *2020 IEEE International Conference on Big Data (Big Data)* (IEEE, 2020), pp. 3123–3132.
13. N. Pinter-Wollman, A. Jelić, N. M. Wells, *Phil. Trans. R. Soc. B* **373**, 20170245 (2018).
14. T. Fisher, *Places J.* 10.22269/101018 (2010).
15. M. Gharipour, A. Tchikine, Eds., *Salutogenic Urbanism: Architecture and Public Health in Early Modern European Cities* (Palgrave Macmillan, 2023).
16. P. Crucitti, V. Latora, S. Porta, *Phys. Rev. E* **73**, 036125 (2006).
17. N. A. Megahed, E. M. Ghoneim, *Sustain Cities Soc.* **61**, 102350 (2020).
18. S. Cremer, S. A. O. Armitage, P. Schmid-Hempel, *Curr. Biol.* **17**, R693–R702 (2007).
19. A. Perna, G. Theraulaz, *J. Exp. Biol.* **220**, 83–91 (2017).
20. N. Stroeymeyt, B. Casillas-Pérez, S. Cremer, *Curr. Opin. Insect Sci.* **5**, 1–15 (2014).
21. S. O'Fallon, E. S. H. Lowell, D. Daniels, N. Pinter-Wollman, *Behav. Ecol.* **33**, 644–653 (2022).
22. N. J. Minter, N. R. Franks, K. A. Brown, *J. R. Soc. Interface* **9**, 586–595 (2012).
23. M. Sankovitz, J. Purcell, *Sci. Rep.* **11**, 23053 (2021).
24. F. García Ibarra, P. Jouquet, N. Bottinelli, A. Bultelle, T. Monnin, *J. Anim. Ecol.* **93**, 319–332 (2024).
25. E. Toffin, J. Kindekens, J.-L. Deneubourg, *Proc. R. Soc. B* **277**, 2617–2625 (2010).
26. J.-B. Leclerc, J. Pinto Silva, C. Detrain, *R. Soc. Open Sci.* **5**, 180267 (2018).
27. C. D. Pull *et al.*, *eLife* **7**, e32073 (2018).
28. A. Reber, M. Chapuisat, *Insectes Soc.* **59**, 231–239 (2012).
29. M. Kim *et al.*, *Bull. World Health Organ.* **98**, 842–848 (2020).
30. P. Rasse, J.-L. Deneubourg, *J. Insect Behav.* **14**, 433–449 (2001).
31. M. Salathé, J. H. Jones, *PLOS Comput. Biol.* **6**, e1000736 (2010).
32. M. L. Mouronte-López, *J. Adv. Transp.* **2021**, 5513311 (2021).
33. E. Dudkina *et al.*, *Int. J. Control* **97**, 1325–1340 (2023).
34. S. M. Kissler *et al.*, *Epidemics* **26**, 86–94 (2019).
35. N. Gravish, G. Gold, A. Zangwill, M. A. Goodisman, D. I. Goldman, *Soft Matter* **11**, 6552–6561 (2015).
36. J. Chang, S. Powell, E. J. H. Robinson, M. C. Donaldson-Matasci, *Swarm Intell.* **15**, 7–30 (2021).
37. T. O. Richardson, N. Stroeymeyt, A. Crespi, L. Keller, *Nat. Commun.* **13**, 6985 (2022).
38. S. Stockmaier *et al.*, *Science* **371**, eabc8881 (2021).
39. A.-L. Barabási, R. Albert, *Science* **286**, 509–512 (1999).
40. N. J. A. Sloane, S. Plouffe, *The Encyclopedia of Integer Sequences* (Academic Press, 1995).
41. M. R. Pie, R. B. Rosengaus, J. F. Traniello, *J. Theor. Biol.* **226**, 45–51 (2004).
42. Z. Cook, D. W. Franks, E. J. H. Robinson, *Behav. Ecol. Sociobiol.* **68**, 509–517 (2014).
43. E. V. Grizanova, C. J. Coates, I. M. Dubovskiy, T. M. Butt, *Virulence* **10**, 999–1012 (2019).
44. J. J. Baljon, J. T. Wilson, *Curr. Opin. Immunol.* **77**, 102215 (2022).
45. V. Latora, M. Marchiori, *Phys. Rev. Lett.* **87**, 198701 (2001).
46. M. Barthélemy, A. Barrat, R. Pastor-Satorras, A. Vespignani, *J. Theor. Biol.* **235**, 275–288 (2005).
47. I. Z. Kiss, D. M. Green, R. R. Kao, *J. R. Soc. Interface* **3**, 669–677 (2006).
48. P. Sah, S. T. Leu, P. C. Cross, P. J. Hudson, S. Bansal, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4165–4170 (2017).
49. E. M. Volz, J. C. Miller, A. Galvani, L. Ancel Meyers, *PLOS Comput. Biol.* **7**, e1002042 (2011).
50. M. E. J. Newman, *Soc. Networks* **27**, 39–54 (2005).
51. S. Wang, Y. Du, Y. Deng, *Commun. Nonlinear Sci. Numer. Simul.* **47**, 151–163 (2017).
52. L. Leckie, M. S. Andon, K. Bruce, N. Stroeymeyt, Architectural immunity: Ants alter their nest networks to prevent epidemics, Zenodo (2025). https://doi.org/10.5281/zenodo.15353580

ACKNOWLEDGMENTS

We thank L. Martin-Silverstone and the XTM Facility, Palaeobiology Research Group, University of Bristol, for use and training for micro-CT scanning. We thank N. V. Meyling from the University of Copenhagen for providing the *M. brunneum* strain. We also thank A. Perna for advice on how to extract networks from 3D image data and R. Brown, F. Masson, T. Richardson, D. Schläppi, and A. Wanderlingh for comments on the manuscript. **Funding:** This work was supported by the European Research Council (ERC starting grant DISEASE, no. 802628, to N.S.). **Author contributions:** L.L. and N.S. designed the study. M.S.A. and K.B. designed and performed the video analysis. M.S.A., K.B., and L.L. conducted the experiment. L.L. designed the code for the network extraction, agent-based model, analysis, and figure generation, with input from N.S. L.L. and N.S. wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data are available at Zenodo (*52*). All code is available through Github (https://github.com/Luke-Leckie/Architectural_Immunity_Nest_Networks).

License information: Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse.

This research was funded in whole or in part by the European Research Council (grant no. 802628); as required the author will make the Author Accepted Manuscript (AAM) version available under a CC BY public copyright license.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ads5930
Materials and Methods; Supplementary Text; Figs. S1 to S11; Tables S1 to S4; References (*53*–*64*); Movie S1

Submitted 20 August 2024; resubmitted 23 January 2025; accepted 12 August 2025
10.1126/science.ads5930

ORGANIC CHEMISTRY

Visible light–driven stereodivergent allylation of cyclic hemiacetals with butene for polypropionate synthesis

Hiroyasu Nakao[1]†, Mirja Md Mahamudul Hassan[1]†, Yusuke Nakamura[1], Moe Toyobe[1], Masahiro Higashi[2], Harunobu Mitsunuma[1]*, Motomu Kanai[1]*

Catalytically transforming abundant hydrocarbon feedstocks into structurally complex, high-value molecules is a pivotal yet challenging goal in organic synthesis. The key difficulty lies in the simultaneous activation of chemically inert feedstocks and precise stereochemical control. Here, we report a catalytic stereodivergent allylation of unprotected cyclic hemiacetal aldols with butene, enabling the programmable synthesis of polypropionates—privileged structural motifs prevalent in biologically active compounds, including pharmaceuticals. This visible light–driven, selective transformation exhibits broad functional group compatibility, furnishing 1,3-polyols with multiple contiguous stereocenters in high yield and stereochemical fidelity. Moreover, this method provides a concise and practical route to key natural product intermediates with minimal protection–deprotection sequences. This strategy has the potential to streamline polypropionate synthesis while reducing the time, cost, and environmental impact.

Organic synthesis plays a crucial role in the global production and supply of high-value molecules, such as pharmaceuticals and functional materials, essential for societal well-being. In the context of 21st-century organic synthesis, there is a growing demand for efficient methodologies to synthesize fine chemicals from readily available precursors in clean and concise ways. In particular, using abundant petroleum-derived feedstocks, which represent the most upstream carbon resources in industrial production, to perform precise organic synthesis is a highly desirable approach. The polypropionate motif, a common substructure in naturally occurring polyketides, frequently exhibits considerable bioactivity and is prevalent in small-molecule drugs (*1*, *2*). Because the consecutive stereocenters are densely arrayed within the polypropionate framework, precise stereochemical control of the iterative carbon elongation strategy is required.

Historically, polypropionate synthesis has driven major advancements in synthetic methodology, leading to the development of numerous prominent approaches. However, some hurdles remain unsolved. Sequential aldol reactions and allylation reactions of aldehydes have been regarded as among the most reliable approaches (*3*–*12*) (Fig. 1A). A multitude of stereoselective carbon elongation methodologies have been studied, spearheaded by landmark reactions such as the Evans aldol reaction (*13*) and the allylation reactions using chiral allylboron reagents established by Hoffmann (*3*), Brown (*14*), Roush (*15*), and Masamune (*16*) as well as chiral allylsilicon reagents developed by Panek (*17*) and Leighton (*9*). Recently, Aggarwal and colleagues have achieved considerable success in the stereoselective assembly line synthesis of polypropionates through a lithiation-borylation strategy (*18*, *19*). However, these methods often necessitate multiple steps to prepare the requisite carbon units, which can pose challenges in terms of overall step economy and waste management. Krische and coworkers have pioneered a catalytic asymmetric hydrogen transfer–based crotylation using alcohols and butadiene (*20*, *21*) (Fig. 1A). This approach enables single-step carbon chain elongation using petroleum-derived starting materials. They have elegantly demonstrated polypropionate synthesis through sequential catalytic crotylation (*22*, *23*). However, in iterative reactions, the stereochemical mismatches between stereocenters of the substrate and catalyst posed challenges for stereodivergent synthesis (*14*, *24*). Furthermore, existing methodologies struggle with carbon elongation in the presence of free hydroxy groups, necessitating protection steps during each elongation process, thereby rendering the synthesis less straightforward (*25*). This issue arises not only from the protonolysis and deactivation of the carbon nucleophiles but also from the formation of cyclic hemiacetals from δ-hydroxy aldehydes, which halts further elongation (*26*).

Given the aforementioned context, proline- and Cu(I)-catalyzed iterative asymmetric aldol reactions have been developed to construct polypropionate frameworks with varied stereochemical configurations, particularly up to the double aldol cyclic hemiacetal stage (*5*, *26*–*29*) (Fig. 1B). However, these approaches face fundamental limitations in further carbon chain elongation. For instance, the proline-catalyzed aldol reactions of propanal to cyclic hemiacetals exhibit poor substrate generality and limited reactivity (*30*). Moreover, the chiral Cu(I)-catalyzed aldol reaction exhibits no reactivity toward cyclic hemiacetal substrates (fig. S1). The latter catalytic system has been restricted to triple additions of identical enolates to simple aldehydes, yielding products with a uniform all-syn stereochemical pattern (*26*) (Fig. 1B). These constraints diminish the synthetic versatility and broader applicability of such methods. As a result, although the concept of stepwise elongation of carbon chains from unprotected double aldol cyclic hemiacetals remains promising for polypropionate synthesis, no existing methodology has achieved both high reactivity and flexible stereochemical control in an operationally simple fashion.

In this study, we developed the streamlined synthesis of polypropionates by catalytic stereoselective crotylation using butene, a petroleum-derived feedstock, as a carbon elongation unit (Fig. 1C). This reaction uses previously inert cyclic hemiacetal aldols as a starting material, avoiding the need for preprotection of hydroxy groups. Furthermore, the catalyst's stereoconfiguration, not the substrates' stereochemical bias, predominantly controls the products' stereochemical outcomes, enabling stereodivergent carbon chain elongation. This methodology has enabled efficient synthesis of natural product core skeletons bearing the polypropionate structure.

Reaction development

Nonprotected cyclic hemiacetal double aldols can be readily synthesized through redox-neutral catalytic asymmetric cross-aldol reactions between aldehydes (*5*, *26*–*29*). To realize a catalytic crotylation of cyclic hemiacetal aldols, several prerequisites must be met. First, it is essential to conduct the reaction in the presence of hydroxy groups, while preventing deactivation of allyl metal species by coordination or protonolysis. Second, given the low concentration of the transiently formed ring-opened aldehydes from cyclic hemiacetals, highly active allyl metal species are required (*31*, *32*). Third, the reaction necessitates catalytic stereochemical control. To fulfill these requirements, our attention was directed toward a ternary catalyst system developed by our group, comprising a photoredox catalyst, a hydrogen atom transfer (HAT) catalyst, and a chromium complex catalyst (Fig. 2A) (*33*). This reaction proceeds through allylchromium species (*34*–*37*), catalytically generated through HAT from an allylic C–H bond of simple alkenes and oxidative interception of the resulting allyl radical

[1]Graduate School of Pharmaceutical Sciences, The University of Tokyo, Tokyo, Japan. [2]Graduate School of Informatics, Nagoya University, Nagoya, Japan. *Corresponding author. Email: h-mitsunuma@mol.f.u-tokyo.ac.jp (H.M.); kanai@mol.f.u-tokyo.ac.jp (M.K.)
†These authors contributed equally to this work.

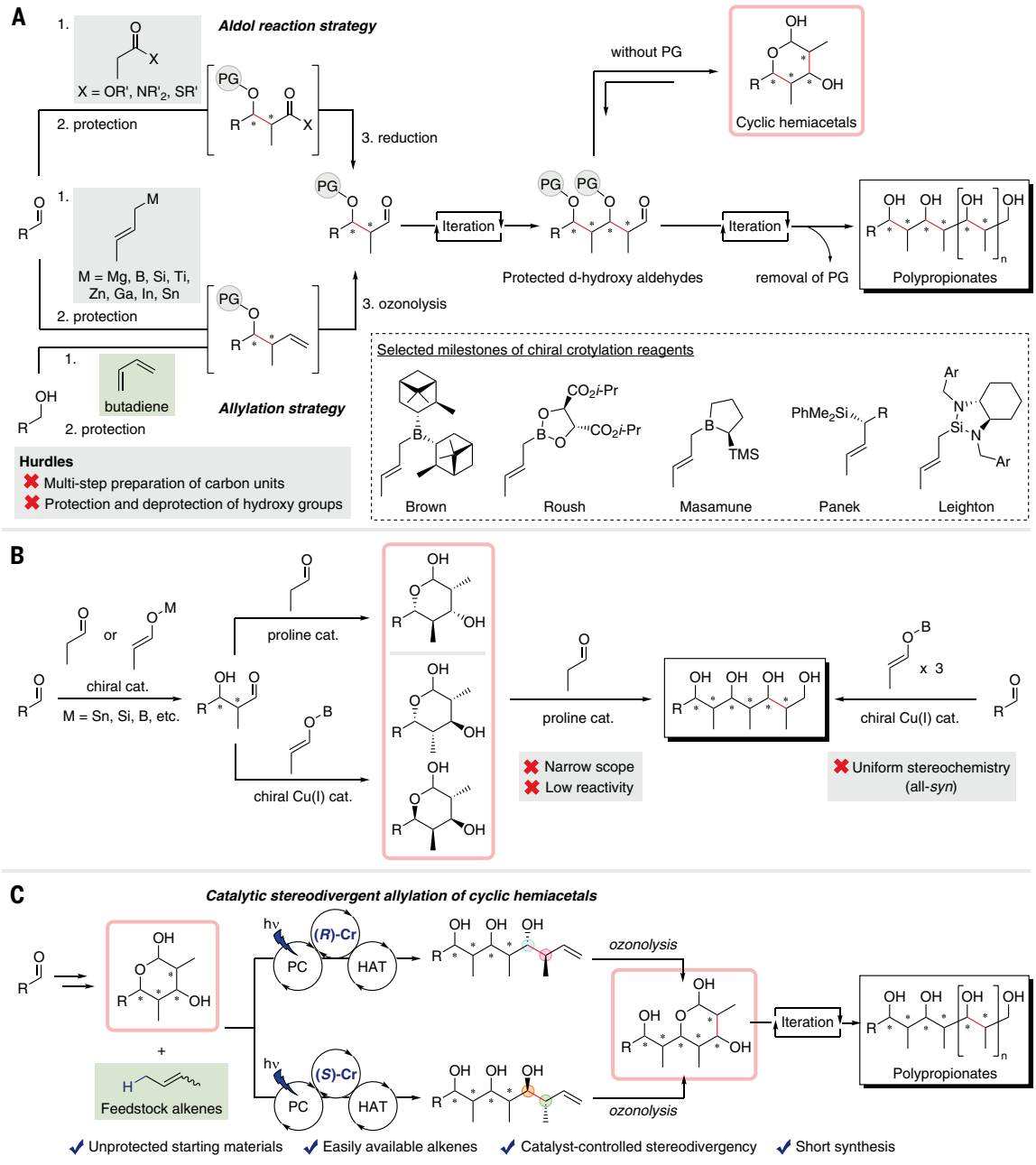
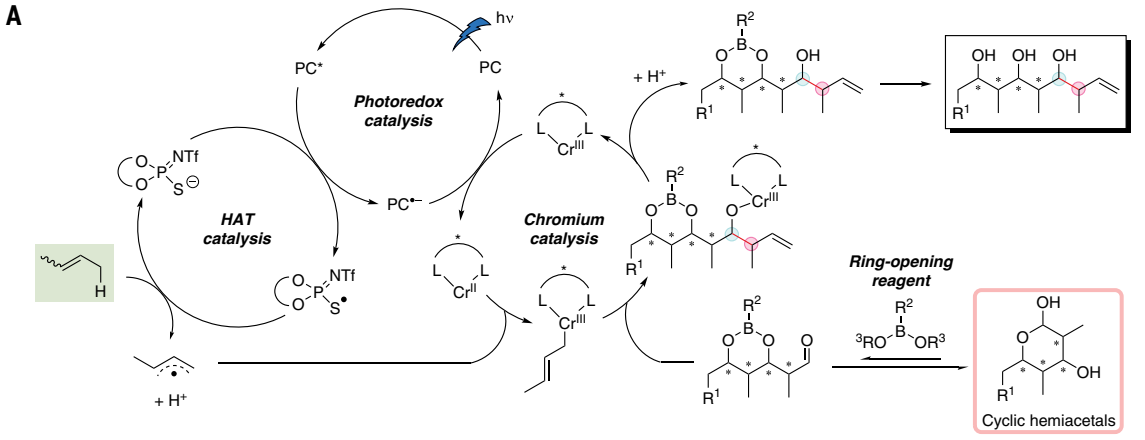
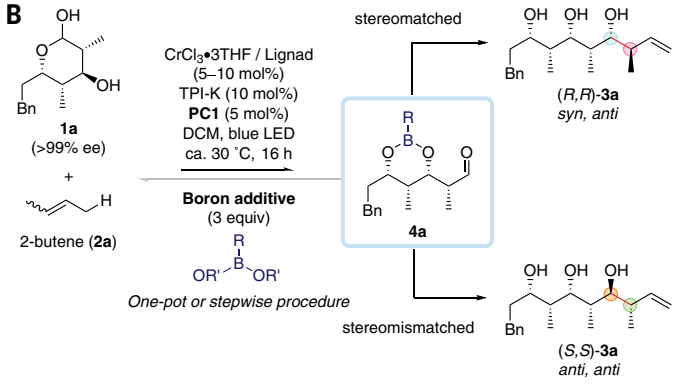


Fig. 1. Iterative stereocontrolled synthesis of a polypropionate motif. (**A**) Previous iterative synthesis of polypropionates. PG, protecting group. (**B**) Previous synthesis of a tripropionate unit via catalytic asymmetric aldol reaction of aldehydes without the use of protecting groups. (**C**) This work. PC, photoredox catalyst.

by a chiral chromium complex. The reaction exhibits high functional group compatibility and is applicable to asymmetric transformations.

Consequently, we initiated optimization of the reaction conditions using cyclic hemiacetal **1a** [>99% enantiomeric excess (ee)] (*26*) with 2-butene (**2a**: E/Z mixture) (Fig. 2B; see table S1 for details). The previous optimal conditions, involving acridinium photoredox catalyst **PC1**, TPI-K HAT catalyst, and achiral $CrCl_3$•*3*THF complex, resulted in no crotylation reaction (Fig. 2B, entry 1). This low reactivity was likely due to the minimal concentration of aldehyde present in the system. To promote the ring-opening of **1a**, we studied the effects of a boron reagent (*3I*) (Fig. 2A). The addition of trimethyl borate (3 equiv) to the reaction mixture promoted the formation of product **3a** in 3% yield (Fig. 2B, entry 2). Upon examining various boron reagents (Fig. 2B, entries 2 to 5), we found that $MeB(Oi\text{-}Pr)_2$ provided (*R*,*R*)-**3a** in the highest yield of 50% with a 12/1 diastereomeric ratio (dr) (Fig. 2B, entry 5). NMR experiments revealed that, in the presence of hemiacetal **1a** and $MeB(Oi\text{-}Pr)_2$, 17% of the ring-opened species, possibly **4a**, was observed, whereas <1% ring-opening was observed in the absence of the boron additive. This suggests that the reaction proceeds through aldehyde **4a**, generated reversibly in the reaction mixture (fig. S2). Density functional theory calculations supported that this ring-opening process proceeds efficiently at ambient temperature; ring-opened aldehyde **4a** was 0.7 kcal/mol more stable than **1a**, with an activation energy of 20.8 kcal/mol (figs. S3 and S4). Introducing a chiral ligand, (*R*)-**L1**, increased the yield and diastereoselectivity to 90% with 14/1 dr (78% isolated yield; Fig. 2B, entry 6: condition A). This reaction required the presence of a chromium complex catalyst, a HAT catalyst, a photoredox catalyst, and light irradiation; it did not proceed if any of these components were



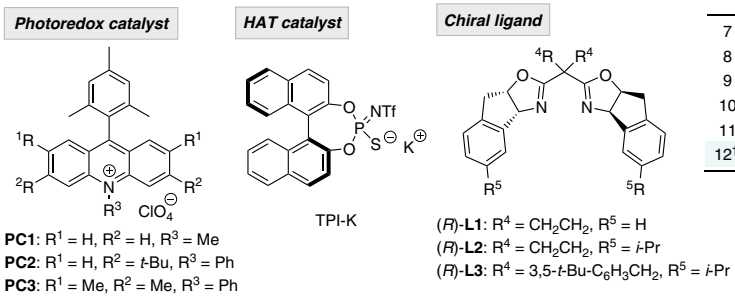


Optimization of reaction conditions for (*R,R*)-3a

entry	Boron additive	Ligand	yield (%)	dr
1	none	none	0	ND
2	$B(OMe)_3$	none	3	ND
3	$B(OEt)_3$	none	25	9.4/1
4	$B(O\textit{i}\text{-}Pr)_3$	none	12	8.3/1
5	$MeB(O\textit{i}\text{-}Pr)_2$	none	50	12/1
6	$MeB(O\textit{i}\text{-}Pr)_2$	(*R*)-**L1**	90 (78)*	14/1

Optimization of reaction conditions for (*S,S*)-3a

entry	Boron additive	Ligand	yield (%)	dr
7	$MeB(O\textit{i}Pr)_2$	(*S*)-**L1**	25	1.5/1
8	$PhB(OH)_2$	(*S*)-**L1**	53	2.0/1
9	$p\text{-}CF_3\text{-}C_6H_4B(OH)_2$	(*S*)-**L1**	71	2.0/1
10	$p\text{-}CF_3\text{-}C_6H_4B(OH)_2$	(*S*)-**L2**	76	2.5/1
11	$p\text{-}CF_3\text{-}C_6H_4B(OH)_2$	(*S*)-**L3**	37	6.0/1
12†	$p\text{-}CF_3\text{-}C_6H_4B(OH)_2$	(*S*)-**L3**	70 (58)*	9.0/1



PC1: R^1 = H, R^2 = H, R^3 = Me
PC2: R^1 = H, R^2 = *t*-Bu, R^3 = Ph
PC3: R^1 = Me, R^2 = Me, R^3 = Ph

(*R*)-**L1**: R^4 = CH_2CH_2, R^5 = H
(*R*)-**L2**: R^4 = CH_2CH_2, R^5 = *i*-Pr
(*R*)-**L3**: R^4 = 3,5-*t*-Bu-$C_6H_3CH_2$, R^5 = *i*-Pr

Fig. 2. Optimization of reaction conditions. (**A**) Proposed catalytic cycle. (**B**) Optimization of reaction conditions for the synthesis of (*R,R*)-**3a** and (*S,S*)-**3a**. Yields and diastereomeric ratios were determined by proton nuclear magnetic resonance (^{1}H NMR) with reference to an internal standard (see the supplementary materials for details). Optimized conditions for (*R,R*)-**3a** (condition A: entry 6): **1a** (1 equiv), **2a** (E/Z mixture, 56 equiv), **PC1** (5 mol %), TPI-K (10 mol %), $CrCl_3$•3THF/(*R*)-**L1** (10/10 mol %), and $MeB(O\textit{i}\text{-}Pr)_2$ (3 equiv) were combined in dichloromethane (DCM; 0.1 M) at room temperature under blue LED irradiation for 16 hours. Optimized conditions for (*S,S*)-**3a** (condition B: entry 12): **1a** (1 equiv), $p\text{-}CF_3\text{-}C_6H_4B(OH)_2$ (1.05 equiv), and *i*-PrOH (15 equiv) were combined in dichloroethane (DCE; 0.2 M) at 150°C for 14 hours. After evaporation, the crude mixture was treated with **2a** (E/Z mixture, 56 equiv), **PC1** (5 mol %), TPI-K (10 mol %), $CrCl_3$•3THF/(*S*)-**L3** (5/7.5 mol %), and $MgCl_2$ (1 equiv) in DCM (0.1 M) at room temperature under blue LED irradiation for 13 hours. TPI-K, thiophosphoric imide potassium salt. *Isolated yield. †$MgCl_2$ (1 equiv) was added.

missing (table S1). The allylation reaction of cyclic hemiacetal **1a** did not proceed (*37*–*39*) or proceeded with low diastereoselectivity (*32*) under existing conditions (fig. S5), indicating that the advantage of the current catalytic system lies in its high functional group compatibility and stereoselectivity with a HAT mechanism.

Next, we investigated whether this system could be extended to catalyst-controlled stereodivergent reactions (*40*). Controlling stereochemistry to achieve high selectivity for the anti,anti-configuration in aldehyde crotylation reactions has been a long-standing challenge, with effective reactions remaining limited even using stoichiometric chiral auxiliaries (*24*, *41*). Using ligand (*S*)-**L1**, which simply inverts the configuration, resulted in a stereochemical mismatch between the chiral allylchromium species and the substrate, leading to low yield and diastereoselectivity (25% yield and 1.5/1 dr; Fig. 2B, entry 7). A stepwise procedure using $p\text{-}CF_3\text{-}C_6H_4B(OH)_2$ at 150°C for the ring-opening followed by catalytic crotylation enhanced the yield to 71% with a 2.0/1 dr (entry 9; see table S2 for optimization of the ring-opening conditions). We then investigated various chiral ligands (fig. S6) and found that the diastereoselectivity was improved to 6.0/1 using ligand (*S*)-**L3** containing isopropyl groups at the phenyl moieties and sterically bulky benzyl groups at the linkage of the bisoxazoline moieties; however, yield decreased to 37% (entry 11). The addition of $MgCl_2$ facilitated the reaction to yield the product (*S,S*)-**3a** in 70% with 9.0/1 dr (58% isolated yield; Fig. 2B, entry 12: condition B; table S3).

Substrate scope exploration

On the basis of the optimized reaction conditions, the substrate scope was investigated (Fig. 3). This stereodivergent crotylation proceeded efficiently

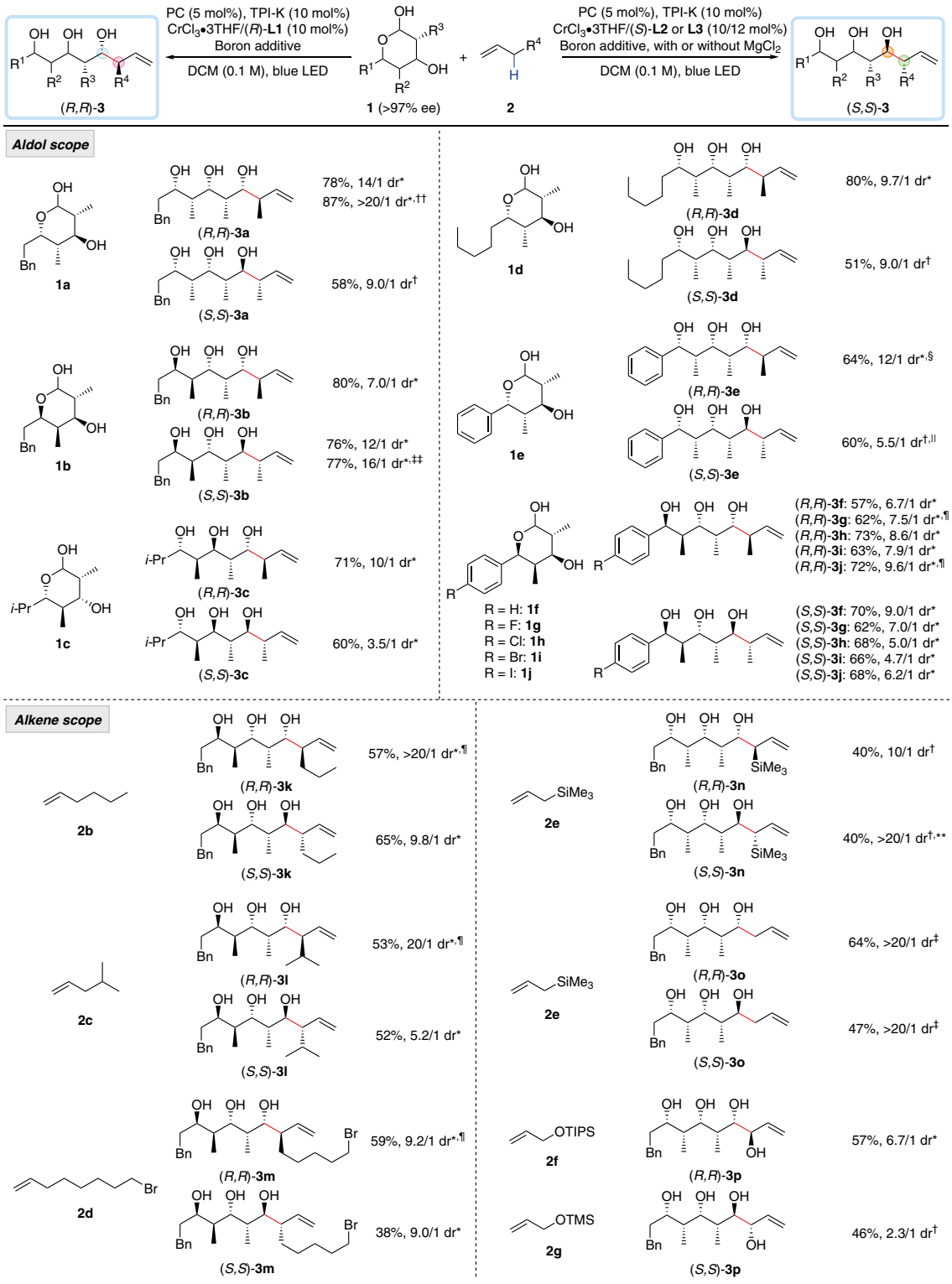


Fig. 3. Substrate scope. Yields are isolated yields. Diastereomeric ratios were determined by ^{1}H NMR analysis of crude mixtures. *Condition A: **1** (1 equiv), **2** (>5 equiv), **PC1** (5 mol %), TPI-K (10 mol %), $CrCl_3$•3THF/(*R*)-**L1** (10/10 mol %) or $CrCl_3$•3THF/(*S*)-**L2** (10/12 mol %), and $MeB(Oi\text{-}Pr)_2$ (3 equiv) were reacted in dichloromethane (DCM; 0.1 M) at room temperature under blue LED irradiation for 16 hours. †Condition B: **1** (1 equiv), *p*-CF_3-$C_6H_4B(OH)_2$ (1.05 equiv), and *i*-PrOH (15 equiv) were reacted in dichloroethane (DCE; 0.2 M) at 150°C for 14 hours. After evaporation, the residue was reacted with **2** (>5 equiv), **PC1** (5 mol %), TPI-K (10 mol %), and $CrCl_3$•3THF/(*R*)-**L1** (10/10 mol %) or $CrCl_3$•3THF/(*S*)-**L3** (5/7.5 mol %) in DCM (0.1 M) at room temperature under blue LED irradiation for 13 hours. ‡Condition C: **1** (1 equiv), *p*-CF_3-$C_6H_4B(OH)_2$ (1.05 equiv), and *i*-PrOH (15 equiv) were reacted in DCE (0.2 M) at 150°C for 14 hours. After evaporation, the residue was reacted with **2e** (5 equiv), **PC2** (5 mol %), and $CrCl_3$•3THF /(*R*)-**L1** (10/10 mol %) or $CrCl_3$•3THF/(*S*)-**L1** (20/24 mol %) in DCM (0.1 M) at room temperature under blue LED irradiation for 24 hours. §$CrCl_3$•3THF/(*R*)-**L1** (15/15 mol %) was used. ||$CrCl_3$•3THF/(*S*)-**L3** (10/12 mol %) was used. ¶**PC2** (5 mol %) was used. **TPI-K (30 mol %) and $CrCl_3$•3THF/(*S*)-**L1** (10/12 mol %) were used. ††Reaction was conducted on a 1-g scale. ‡‡ Reaction was conducted on a 500-mg scale. See the supplementary materials for details.

with cyclic hemiacetals **1** bearing various stereochemical configurations, which were accessible through proline- or Cu(I)-catalyzed double aldol reaction (*5*, *26*). For substrates such as **1b** and **1c**, the stereomismatched crotylation proceeded in the presence of $MeB(O\textit{i}\text{-}Pr)_2$ without requiring prior boron-mediated ring-opening [condition A, (*S*)-**L2** was used instead of (*R*)-**L1**; see fig. S7]. The susceptibility to ring-opening, depending on the stereochemistry, is likely to determine whether condition A or B was suitable. The reaction was also applicable to cyclic hemiacetals derived from various aliphatic and aromatic aldehydes (**1d** to **1j**). The applicability of the reaction was subsequently extended to a range of alkene substrates. The reaction proceeded smoothly when long-chain alkene **2b** or branched alkene **2c** was used. Bromine-substituted alkene **2d** was also competent. When allylsilane **2e** was used, the reaction yielded silicon-substituted products **3n**. Additionally, the Hosomi-Sakurai–type allylation reaction proceeded, affording silicon-nonsubstituted products **3o**, without the TPI-K HAT catalyst (condition C). In this case, allylchromium species were generated through the single-electron oxidation of **2e** by photoexcited **PC2**, followed by dissociation of the TMS group to generate allyl radical, which was intercepted by the chiral chromium complex catalyst (*42*). This demonstrated the feasibility of introducing three carbon units, which are also critical in polyketide natural product synthesis. Furthermore, the reaction proceeded efficiently with protected allyl alcohols **2f** or **2g**, enabling the stereodivergent synthesis of tetraols **3p** containing a 1,2-diol motif. In general, the reaction proceeded with high diastereoselectivity; however, the diastereomeric ratios of (*S*,*S*)-**3c** and (*S*,*S*)-**3p** were limited to 3.5/1 and 2.3/1, respectively. This method is amenable to scale-up, proceeding without any loss in reactivity or selectivity: 87% yield with >20/1 dr for producing (*R*,*R*)-**3a** on a 1-g scale and 77% yield with 16/1 dr for producing (*S*,*S*)-**3b** on a 500-mg scale (Fig. 3). Furthermore, we could recover 49% of $p\text{-}CF_3\text{-}C_6H_4B(OH)_2$ after the reaction between **1a** and **2a** to produce (*S*,*S*)-**3a** under condition B (see section 10 of the supplementary materials). Thus, this method holds considerable promise for broad adoption in practical synthesis.

This crotylation proceeds under mild conditions and is characterized by its high functional group compatibility. Leveraging these advantages, the reaction was applied to stereodivergent crotylation of structurally complex substrates (Fig. 4A). The reaction proceeded with cholesterol derivative **1k** containing a C=C double bond, yielding the desired crotylation products (*R*,*R*)-**3q** and (*S*,*S*)-**3q** stereodivergently. Similarly, the reaction with peptide derivative **1l**, bearing multiple coordinating functional groups, afforded products **3r** with high selectivity. Furthermore, the crotylated products are versatile intermediates for various transformations (Fig. 4B). Through ozonolysis, product **3a** can be converted into cyclic hemiacetal triple aldol **5a**. This transformation enables its application in subsequent crotylation (see below). Moreover, a diastereoselective diborylation-oxidation sequence (7.9/1 dr) (*43*) enabled the synthesis of enantiomerically and diastereomerically enriched pentol **6**. Convergent coupling with a thiosugar substrate was effectively achieved to produce **7** through a thiol-ene reaction (*44*). Heck reaction allowed for the extension of the carbon skeleton at the terminal alkene site, affording **8**.

Next, we applied this method to further elongation of the carbon chains through iterative processes (Fig. 5A). Following the conversion of products (*R*,*R*)-**3a** and (*S*,*S*)-**3a** to cyclic hemiacetal triple aldols **5a** and **5b**, respectively, through ozonolysis, we investigated the catalytic crotylation reaction. The ring-opening efficiency decreased compared with double aldols **1**, necessitating a slight increase in the reaction temperature (44°C). For the reaction of **5a** to **9a**, yield improved from 8% to 52% by the addition of 10 equiv *i*-PrOH (isopropyl alcohol; table S4). The alcohol likely facilitates the ring-opening process by accelerating the interconversion between regioisomeric boron acetals. The stereomatched crotylation reaction proceeded smoothly to produce quadruple aldol equivalents **9a** and **9b**. The generality of the crotylation reaction for the cyclic hemiacetal triple aldols **5** was examined (Fig. 5B). The reaction proceeded with moderate to high yields across a variety

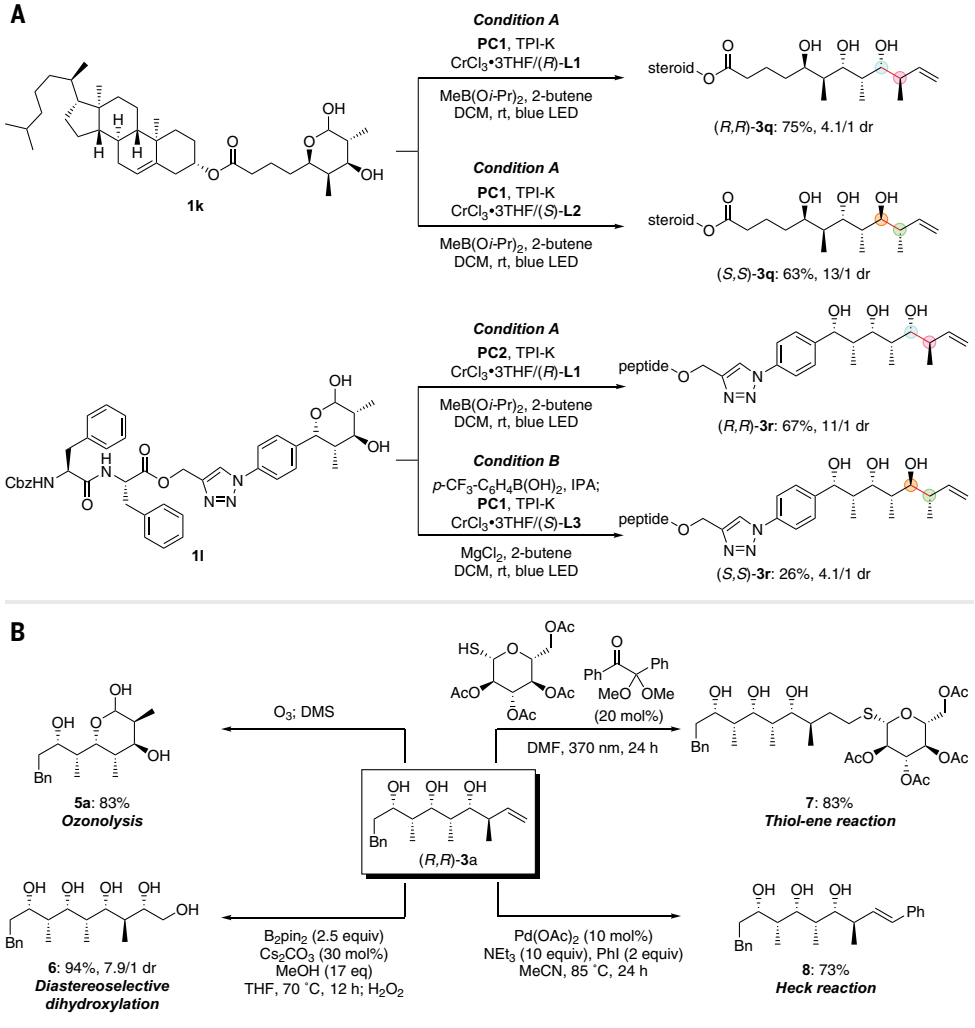


Fig. 4. Late-stage applications and synthetically useful product transformations. (**A**) Catalytic stereodivergent crotylation using multifunctional molecules. Yields are isolated yields. Diastereomeric ratios were determined by ^{1}H NMR analysis of crude mixtures. For the reaction conditions, see the footnote of Fig. 3 and the supplementary materials. (**B**) Transformations of product (*R*,*R*)-**3a**.

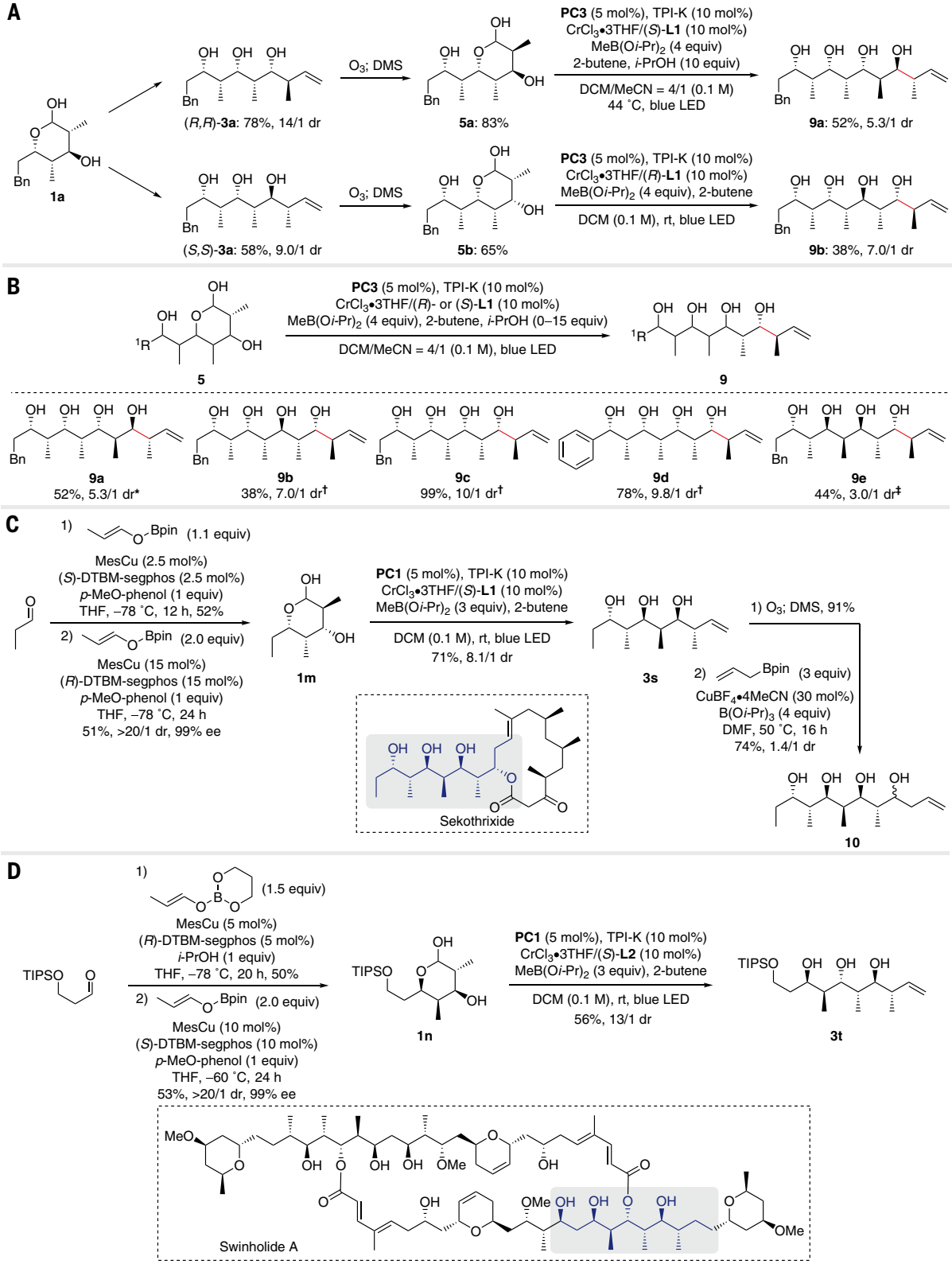


Fig. 5. Iterative processes. (**A**) Iterative crotylation from **1a**. (**B**) Substrate scope of the catalytic crotylation of cyclic hemiacetal triple aldols **5**. Yields are isolated yields. Diastereomeric ratios were determined by ^{1}H NMR analysis of crude mixtures. *__5__ (1 equiv), 2-butene (**2a**, E/Z mixture, 56 equiv), **PC3** (5 mol %), TPI-K (10 mol %), $CrCl_3$•3THF/(*S*)-**L1** (10 mol %), MeB(O*i*-Pr)$_2$ (4 equiv), and *i*-PrOH (10 equiv) were combined in DCM/MeCN = 4/1 (0.1 M) at 44°C under blue LED irradiation for 48 hours. †__5__ (1 equiv), 2-butene (**2a**, E/Z mixture, 56 equiv), **PC3** (5 mol %), TPI-K (10 mol %), $CrCl_3$•3THF/(*R*)-**L1** (10 mol %), and MeB(O*i*-Pr)$_2$ (4 equiv) were combined in DCM (0.1 M) at room temperature under blue LED irradiation for 48 hours. ‡__5__ (1 equiv), 2-butene (**2a**, E/Z mixture, 56 equiv), **PC3** (5 mol %), TPI-K (10 mol %), $CrCl_3$•3THF/(*R*)-**L1** (10 mol %), MeB(O*i*-Pr)$_2$ (4 equiv), and *i*-PrOH (15 equiv) were combined in DCM/MeCN = 4/1 (0.1 M) at 44°C under blue LED irradiation for 48 hours. (**C**) Application to the synthesis of sekothrixide core structure. (**D**) Application to the synthesis of swinholide A core structure.

of substrates with different substitution patterns (**9a** to **9e**). This result highlights the potential of this catalytic system for iterative reactions.

Finally, this method was applied to the streamlined synthesis of key scaffolds in biologically active natural products (Fig. 5C). Sekothrixide contains a polypropionate motif with consecutive asymmetric carbon centers (*45*). Conventional methods required 23 steps to construct the polypropionate substructure, with an overall yield of 14% (*46*). The corresponding substructure was accessible in only five steps using our method. Thus, enantio- and diastereo-pure cyclic hemiacetal **1m** was synthesized in two steps through the iteration of asymmetric Cu(I)-catalyzed cross-aldol reactions between aldehydes (*26*). Subsequent crotylation with 2-butene yielded a diastereoselective homoallylic alcohol **3s** in 71% yield with an 8.1/1 dr. After purification, ozonolysis, followed by the addition of allyl boronate using a Cu(I) catalyst (*47*), furnished tetraol **10** exhibiting the subunit of secothrixide. Furthermore, the same strategy was successfully used to synthesize the polypropionate framework of swinholide A (*48*, *49*). By leveraging the stereomismatched crotylation reaction of cyclic hemiacetal **1n**, the challenging anti-anti structural motif **3t** was catalytically synthesized with exceptional efficiency.

In summary, we have developed a catalytic methodology for constructing diverse polypropionate frameworks using butene, a feedstock resource, and unprotected cyclic hemiacetal aldols. The reaction was enabled by the use of a ternary catalyst system comprising a photoredox catalyst, thiophosphoric imide HAT catalyst, and chiral chromium complex catalyst along with the addition of a boron reagent as a dynamic ring-opening agent. This approach demonstrated broad applicability to various cyclic hemiacetal double and triple aldols and achieved catalyst-programmed stereodivergent crotylation reaction. Furthermore, the method enabled the efficient synthesis of key polyketide frameworks, paving the way for the streamlined production of drug leads in the future.

REFERENCES AND NOTES

1. J. Rohr, *Angew. Chem. Int. Ed.* **39**, 2847–2849 (2000).
2. A. M. P. Koskinen, K. Karisalmi, *Chem. Soc. Rev.* **34**, 677–690 (2005).
3. R. W. Hoffmann, *Angew. Chem. Int. Ed.* **26**, 489–503 (1987).
4. H. J. M. Gijsen, C.-H. Wong, *J. Am. Chem. Soc.* **117**, 7585–7591 (1995).
5. J. Casas, M. Engqvist, I. Ibrahem, B. Kaynak, A. Córdova, *Angew. Chem. Int. Ed.* **44**, 1343–1345 (2005).
6. B. Schetter, R. Mahrwald, *Angew. Chem. Int. Ed.* **45**, 7506–7525 (2006).
7. S. E. Denmark, G. L. Beutner, *Angew. Chem. Int. Ed.* **47**, 1560–1638 (2008).
8. J. Li, D. Menche, *Synthesis (Stuttg.)* **2009**, 2293–2315 (2009).
9. H. Kim, S. Ho, J. L. Leighton, *J. Am. Chem. Soc.* **133**, 6517–6520 (2011).
10. W. Gati, H. Yamamoto, *Acc. Chem. Res.* **49**, 1757–1768 (2016).
11. J. Wu, K. L. Verboom, M. J. Krische, *Chem. Rev.* **124**, 13715–13735 (2024).
12. K. Pieper, R. Bleith, C. Köhler, R. Mika, A. Gansäuer, *Angew. Chem. Int. Ed.* **63**, e202317525 (2024).
13. D. A. Evans *et al.*, *Pure Appl. Chem.* **53**, 1109–1127 (1981).
14. H. C. Brown, P. K. Jadhav, *J. Am. Chem. Soc.* **105**, 2092–2093 (1983).
15. W. R. Roush, A. E. Walts, L. K. Hoong, *J. Am. Chem. Soc.* **107**, 8186–8190 (1985).
16. J. Garcia, B. M. Kim, S. Masamune, *J. Org. Chem.* **52**, 4831–4832 (1987).
17. N. F. Jain, N. Takenaka, J. S. Panek, *J. Am. Chem. Soc.* **118**, 12475–12476 (1996).
18. T. Bootwicha, J. M. Feilner, E. L. Myers, V. K. Aggarwal, *Nat. Chem.* **9**, 896–902 (2017).
19. S. G. Aiken *et al.*, *Nat. Chem.* **15**, 248–256 (2023).
20. J. R. Zbieg, E. Yamaguchi, E. L. McInturff, M. J. Krische, *Science* **336**, 324–327 (2012).
21. E. L. McInturff, E. Yamaguchi, M. J. Krische, *J. Am. Chem. Soc.* **134**, 20628–20631 (2012).
22. A.-M. R. Dechert-Schmitt, D. C. Schmitt, X. Gao, T. Itoh, M. J. Krische, *Nat. Prod. Rep.* **31**, 504–513 (2014).
23. J. Feng, Z. A. Kasun, M. J. Krische, *J. Am. Chem. Soc.* **138**, 5467–5478 (2016).
24. M. Chen, W. R. Roush, *J. Am. Chem. Soc.* **134**, 3925–3931 (2012).
25. I. S. Young, P. S. Baran, *Nat. Chem.* **1**, 193–205 (2009).
26. L. Lin *et al.*, *J. Am. Chem. Soc.* **137**, 15418–15421 (2015).
27. S. E. Denmark, S. K. Ghosh, *Angew. Chem. Int. Ed.* **40**, 4759–4762 (2001).
28. A. B. Northrup, D. W. C. MacMillan, *J. Am. Chem. Soc.* **124**, 6798–6799 (2002).
29. T. Amatov *et al.*, *J. Am. Chem. Soc.* **143**, 14475–14481 (2021).
30. Y. Hirao, Y. Kanzaki, H. Mitsunuma, M. Kanai, *Tetrahedron* **98**, 132448 (2021).
31. X.-F. Wei, Y. Shimizu, M. Kanai, *ACS Cent. Sci.* **2**, 21–26 (2016).
32. T. Adak *et al.*, *Nature* **640**, 94–99 (2025).
33. S. Tanabe, H. Mitsunuma, M. Kanai, *J. Am. Chem. Soc.* **142**, 12374–12381 (2020).
34. J. L. Schwarz, F. Schäfers, A. Tlahuext-Aca, L. Lückemeier, F. Glorius, *J. Am. Chem. Soc.* **140**, 12705–12709 (2018).
35. F. Schäfers *et al.*, *ACS Catal.* **10**, 11841–11847 (2020).
36. H. Hu, Z. Shi, X. Guo, F.-H. Zhang, Z. Wang, *J. Am. Chem. Soc.* **146**, 5316–5323 (2024).
37. A. Fürstner, *Chem. Rev.* **99**, 991–1046 (1999).
38. H. Miyaoka, Y. Okubo, *Synlett* **4**, 547–550 (2011).
39. M. Shibasaki, M. Kanai, *Chem. Rev.* **108**, 2853–2873 (2008).
40. S. Krautwald, E. M. Carreira, *J. Am. Chem. Soc.* **139**, 5627–5639 (2017).
41. C. J. Cowden, I. Paterson, *Org. React.* **51**, 1–200 (2004).
42. F. Schäfers, S. Dutta, R. Kleinmans, C. Mück-Lichtenfeld, F. Glorius, *ACS Catal.* **12**, 12281–12290 (2022).
43. T. P. Blaisdell, T. C. Caya, L. Zhang, A. Sanz-Marco, J. P. Morken, *J. Am. Chem. Soc.* **136**, 9264–9267 (2014).
44. M. Fiore, A. Marra, A. Dondoni, *J. Org. Chem.* **74**, 4422–4425 (2009).
45. Y. J. Kim, K. Furihata, A. Shimazu, K. Furihata, H. Seto, *J. Antibiot. (Tokyo)* **44**, 1280–1282 (1991).
46. D. Katsumi *et al.*, *J. Org. Chem.* **84**, 1553–1562 (2019).
47. K. Ishizawa *et al.*, *J. Org. Chem.* **84**, 10615–10628 (2019).
48. S. Carmely, Y. Kashman, *Tetrahedron Lett.* **26**, 511–514 (1985).
49. I. Shin, S. Hong, M. J. Krische, *J. Am. Chem. Soc.* **138**, 14246–14249 (2016).

ACKNOWLEDGMENTS

Funding: This work was supported by JSPS KAKENHI grants JP23H04909 (Green Catalysis Science) (M.K.), JP20H05843 (Dynamic Exciton) and JP23K06045 (H.M.), and JP20H05839 (Dynamic Exciton) (M.H.). **Author contributions:** H.M. and M.K. conceived of and designed the project. H.N., M.M.M.H., Y.N., and M.T. carried out the experiments. M.H. carried out density functional theory calculations. H.M. and M.K. wrote the manuscript. All authors analyzed the data, discussed the results, and proofread the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Crystallographic data are available free of charge from the Cambridge Crystallographic Database Centre (CCDC) under CCDC 2440008 [for (*S,S*)-**3j**]. All additional data are included in the supplementary materials. **License information:** Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adz0686
Materials and Methods; Supplementary Text; Figs. S1 to S7; Tables S1 to S4; References (*50–60*)

Submitted 16 May 2025; accepted 21 August 2025

10.1126/science.adz0686

ANIMAL SYMBIOSIS

Defensive fungal symbiosis on insect hindlegs

Takanori Nishino[1,2]†, Minoru Moriyama[1]*†, Hiromi Mukai[3]†, Masahiko Tanahashi[1,4]†, Takahiro Hosokawa[1,5], Hsin-Yi Chang[4], Shuji Tachikawa[6], Naruo Nikoh[7], Ryuichi Koga[1], Chih-Horng Kuo[4], Takema Fukatsu[1,2,8]*

Dinidorid stinkbugs were reported to possess a conspicuous tympanal organ on female hindlegs. In this study, we show that this organ is specialized to retain microbial symbionts rather than to perceive sound. The organ's surface is not membranous but consists of porous cuticle in which each pore connects to glandular secretory cells. In reproductive females, the hindleg organ is covered with fungal hyphae that grow from the pores. Upon oviposition, the females transfer the fungi from the organ to the eggs, where the hyphae physically protect the eggs against wasp parasitism. The fungi comprise a diversity of mostly low-pathogenicity Cordycipitaceae.

Tympanal organs have repeatedly evolved in diverse insects and were thought to be required for auditory perception (*1*). The Dinidoridae is a small stinkbug family embracing some 100 species representing 16 genera (*2*–*4*). Previous morphological studies reported the presence of a conspicuously enlarged structure on the hind tibiae of adult dinidorid females, which was interpreted as a tympanal organ (*5*), although to our knowledge, no other stinkbug groups possess such female-specific hindleg organs. To address the question of the function of the tympanum, we investigated the Japanese dinidorid stinkbug *Megymenum gracilicorne* and discovered that this stinkbug's hindleg organ is not auditory but a previously unknown type of symbiotic organ.

M. gracilicorne lives on wild cucurbitaceous plants and sometimes infests cultivated cucumbers and pumpkins (*6*) (Fig. 1, A to C). As previously reported for other dinidorid species (*5*), *M. gracilicorne* exhibits a conspicuous sexual dimorphism in the hindlegs. The hind tibiae of adult females are widened and concave in the middle, consisting of an oval area of tympanum, compared with the thin hind tibiae of adult males (Fig. 1, D to H). However, close morphological inspection showed that structural features of the female-specific hindleg organ are atypical of canonical insect tympanal "ears" in that the tympanum was not membranous but composed of a rigid porous cuticle (Fig. 1, H and I). We found that, in reproductively mature females, the tympanum was covered with fibrous white material (Fig. 1, B, C, and J), resembling fungal hyphae luxuriantly growing from the pores (Fig. 1, K and L). Cultivation and microscopic observation of the fibrous material confirmed that it was filamentous fungal hyphae that formed conidia (fig. S1).

Fungus transfer from hindleg organ to eggs

We observed that the gravid females laid eggs in a row, and when each egg was deposited, the females rhythmically scratched the fungus-covered hindleg organ with the tarsal claws of the opposite hindleg and rubbed the egg surface, smearing the fungi onto the eggs (Fig. 1, M to O, and movie S1). Within a few days, the fungal hyphae grew to cover the entire egg mass (Fig. 1, P to R). On hatching, the hyphae attached to the body surface of newborn nymphs (fig. S2 and movie S2), although the fungi were subsequently lost as the nymphs molted and grew.

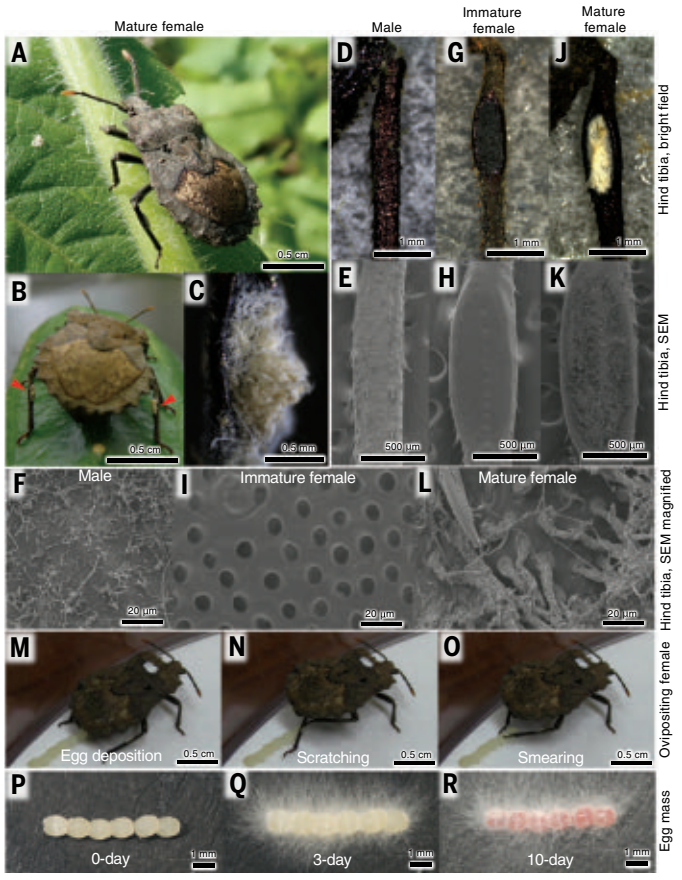


Fig. 1. Morphology and behavior of *M. gracilicorne*. (**A**) An adult female in the field on *Sicyos angulatus*. (**B**) Back view of an adult female reared on cucumber, whose hindleg organs are covered with fungal hyphae (arrowheads). (**C**) Magnified image of the hindleg organ. (**D** to **L**) Morphology of hind tibia of adult male [(D) to (F)], immature adult female [(G) to (I)], and mature adult female [(J) to (L)]. [(D), (G), and (I)] Brightfield images. [(E), (H), and (K)] Scanning electron microscope (SEM) images. [(F), (I), and (L)] Magnified SEM images. (**M** to **O**) Peculiar behavior of egg-laying adult female. After laying each egg (M), the female rhythmically scratches the hyphae-covered hindleg organ with tarsal claws of the opposite hindleg (N) and then rubs the egg surface with the claws in a skillful manner (O). Also see movie S1. (**P** to **R**) Egg mass covered with fungal hyphae at 0-day (P), 3-day (Q), and 10-day (R) intervals after oviposition. Reddish color of 10-day eggs reflects the body color of developing nymphs inside.

Diversity and specificity of symbiotic fungi

The fungi growing on the hind tibiae, hindleg organs, and the eggs were examined for adult females collected from different localities (Fig. 2). Each female was reared in isolation and allowed to lay two egg masses consecutively, which were sampled and subjected to fungal cultivation (data S1 and fig. S3). More than 600 fungal isolates were obtained for sequencing of the ribosomal internal transcribed spacer (ITS) region (data S1). DNA samples extracted from the hind tibiae and the eggs were then subjected to amplicon sequencing analysis for the ITS region (data S2 and fig. S3). Molecular phylogenetic analyses

[1]Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan. [2]Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan. [3]Department of Forest Entomology, Forestry and Forest Products Research Institute, Tsukuba, Japan. [4]Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan. [5]Faculty of Science, Kyushu University, Fukuoka, Japan. [6]Association for Nature Restoration and Conservation, Tokyo, Japan. [7]Department of Liberal Arts, The Open University of Japan, Chiba, Japan. [8]Department of Biological Sciences, Graduate School of Science, the University of Tokyo, Tokyo, Japan. ***Corresponding author. Email: m-moriyama@aist.go.jp (M.M.); t-fukatsu@aist.go.jp (T.F.)** †These authors contributed equally to this work.

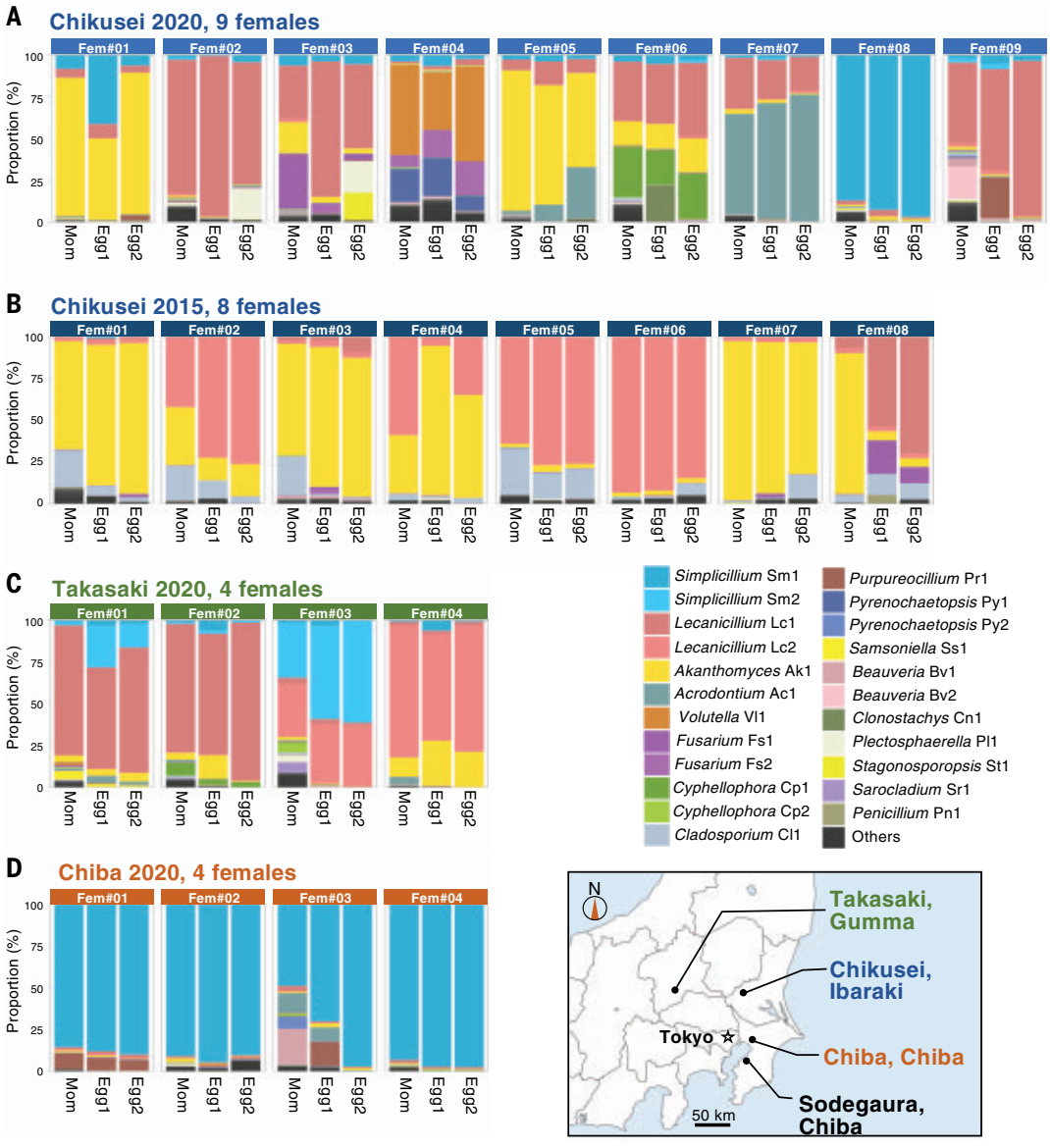


Fig. 2. Fungal microbiota on hindleg organs and eggs of field-collected reproductive adult females of *M. gracilicorne*. (**A**) Nine females collected at Chikusei in 2020. (**B**) Eight females collected at Chikusei in 2015. (**C**) Four females collected at Takasaki in 2020. (**D**) Four females collected at Chiba in 2020. Fungal ITS amplicon sequencing data from the mother's hindlegs (Mom), the first egg mass (Egg1), and the second egg mass (Egg2) are shown by colored bar graphs. For detail, also see fig. S3.

based on these sequence data consistently revealed that the majority of the fungi were Ascomycota, class Sordariomycetes, and family Cordycipitaceae, which include entomopathogenic fungi such as *Cordyceps* and *Beauveria* (*7*, *8*). However, many of the fungi did not belong to highly pathogenic lineages but represented lineages such as *Lecanicillium*, *Simplicillium*, and *Akanthomyces* (*9*–*11*) (figs. S4 and S5).

Fungal diversity and specificity in natural host populations

We found that each female generally hosted multiple fungal species that were similar to the fungal composition of the eggs (Fig. 2 and fig. S3), which appeared to confirm vertical transmission of the fungal associates from the adult female. Notably, within and between collection localities, and between years, fungal composition often markedly differed between individuals (Fig. 2 and fig. S3). These patterns indicated that although *M. gracilicorne* transmits hindleg-associated fungi to eggs vertically, the fungi are lost during nymphal development and have to be newly acquired by adult females from the environment every generation (*12*–*14*).

Structure, cytology, and gene expression of symbionts

The question is how specific fungal strains are collected and selected by the females from the environment. Histological observations showed that, as the females matured, the surface cuticle of the hindleg organ thickened, sclerotized, and became porous (Fig. 3, A to D). No neuronal structures were seen beneath the hindleg organ, but the thick cuticle was lined by an epidermal cell layer (Fig. 3, A to D). Notably, the epidermal cells intruded into the cuticular layer and connected to the bottom of the cuticular pores (Fig. 3, B, and D to G). In mature females, these epidermal columns and the cuticular pores exhibited dense polysaccharide staining (Fig. 3, C and D), indicating secretory activity. Massive fungal proliferation could also be seen in the cuticular pores (Fig. 3, D, E, and H). Comparative transcriptomic analyses of tissues from the midlegs and hindlegs of immature and mature adult insects of *M. gracilicorne*—from which fungus-derived and other contaminating genes were removed (data S3 and S4)—identified 550 genes highly and preferentially expressed in the hindlegs of mature adult females (fig. S6 and data S5 and S6). Functional term enrichment analysis revealed that genes expressing transporters were augmented in the hindleg organ, which is consistent with a secretory function. The most highly expressed gene encoded a Takeout family protein, which is a small secretory protein likely transporting hydrophobic substances (*15*). Genes for UDP-glucuronosyltransferases and the glucuronate pathway were also highly expressed, which presumably were required to mobilize lipophilic materials. Other proteins expressed included a secretory cuticular glycoprotein, possibly involved in the formation of the cuticular structure, some immune-related regulators in the Toll signaling pathway, and others (data S6), although it was unclear whether these genes are involved in fungal selectivity.

Hindleg tympanum as a symbiotic organ for fungi

The female stinkbug's hindleg structure is neither a vibratory membrane nor associated with sensory neurons; instead its function appears to be secretory. The morphological and cytological features of

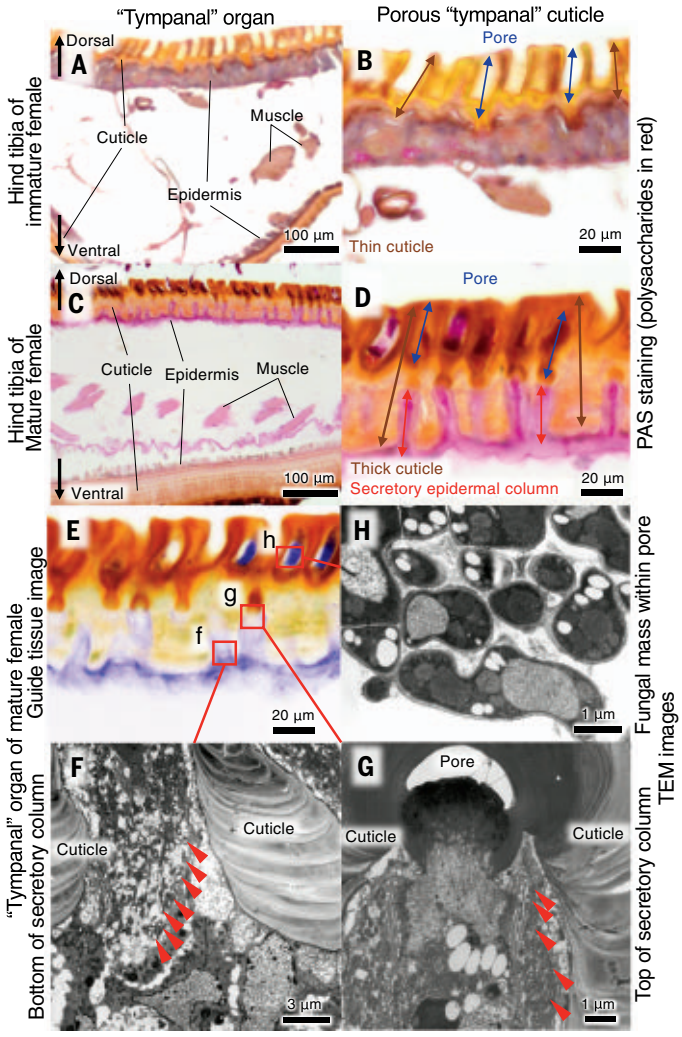


Fig. 3. Histology and fine structure of female-specific hindleg organ of *M. gracilicorne*. (**A** to **D**) Light microscopic images of tissue sections of the hindleg organs subjected to periodic acid–Schiff (PAS) staining and hematoxylin staining, by which sugar-rich elements such as polysaccharides, glycoproteins, mucous secretion, and fungal cell wall are visualized in red. [(A) and (B)] In the immature female, the "tympanal" cuticle is thinner, the epidermal cells are stained in pale pink, and the cuticular pores are empty. [(C) and (D)] In the mature female, the "tympanal" cuticle is thicker and sclerotized, the epidermal cells are deeply stained in red, and the cuticular pores are full of fungal hyphae. No neuronal elements such as chordotonal organs are found beneath the "tympanal" cuticle, and the epidermal cells form secretory columns leading to the bottom of the cuticular pores. (**E**) A light microscopic tissue section image of the "tympanal" cuticle of the mature female for guiding the histological locations, corresponding to transmission electron microscopic (TEM) images (**F** to **H**) shown by red rectangles. (F) TEM image of the bottom of a secretory column. The cytoplasm of secretory epidermal cells full of granules, vacuoles, and membranous elements is seen between thick cuticles. (G) TEM image of the top of a secretory column. The cytoplasm full of secretory granules and membranous elements ends up with the narrow cuticular canals at the bottom of the cuticular pore. Arrowheads highlight secretory granules. (H) TEM image of the content of a cuticular pore; densely packed fungal hyphae are seen.

the pores are reminiscent of the glandular mycangia of ambrosia beetles, which are exoskeletal cavities lined with glandular cells, known for fostering specific symbiotic fungi (*16*, *17*). Similar exoskeletal cavities with glandular structures have also been reported from leaf-cutter ants and digger wasps for hosting symbiotic actinobacteria (*18*, *19*). The large number and high density of the mycangial units concentrated in the specialized region on the hindlegs of *M. graciliforme* are distinctive. Whereas ambrosia beetles generally display a few mycangia (*20*), each hindleg organ of *M. gracilicorne* bears about 2000 fungus-growing pores (mean ± SD = 2118 ± 100; $n = 11$). We conclude that the hindleg organ of this insect species is a previously unknown type of external symbiotic organ consisting of highly concentrated glandular mycangia for hosting specific fungal associates.

Egg surface structure, fungal growth, and proteinaceous secretion

After the maternal fungus-smearing behavior upon oviposition (Fig. 1, M to O, and movie S1), the surface of the eggs became covered with growing fungal hyphae (Fig. 1, P and Q). Microscopy revealed that the egg surface was covered with a polysaccharide-rich layer on which the fungal hyphae appeared to grow (fig. S7, a to f). The polysaccharide-rich layer was already evident on the surface of mature eggs in the female's ovary (fig. S8a) and could be solubilized by detergent with disulfide bond reduction (fig. S8b). The polysaccharide layer was extracted from newly laid eggs (fig. S7, g to k) and was subjected to proteomic analysis (data S7). An abundant secretion protein with an odorant-binding protein (OBP)–like motif was identified. Other proteins that were detected included β-glucuronidase, a putative antibacterial peptide, and ferritin (data S7). Recent work has reported that stinkbugs of the family Plataspidae produce symbiont-containing capsules, in which a predominant secretion protein with an OBP domain provides a protective matrix for the symbiont cells outside the host (*21*). However, the abundant OBP-like motif protein that we found in *M. gracilicorne* was phylogenetically unrelated to the plataspid protein.

Pathogenicity of hindleg-associated fungi

The Cordycipitaceae includes many entomopathogenic fungal species, such as *Cordyceps* and *Beauveria* (*7*, *8*). The fungal associates of *M. gracilicorne* did not appear to be pathogenic to this insect when conidia were applied topically (fig. S9a). When fungal conidia were injected directly into the hemocoel, one species of *Beauveria* promptly killed all the treated insects. By contrast, when injected, the hindleg-derived *Simplicillium* Sm1 and *Lecanicillium* Lc2 caused very little insect mortality; although *Akanthomyces* Ak1 exhibited some pathogenicity, it was less virulent than *Beauveria* (fig. S9b). Possibly, *M. gracilicorne* develops resistance to these potential entomopathogens during the symbiotic association. However, similar experiments with the pentatomid stinkbug *Plautia stali* showed similar responses to the same range of fungi (fig. S9, c and d), and it appears that adult females of *M. gracilicorne* select relatively benign filamentous fungal species.

Symbiosis for pathogen exclusion

In leaf cutter ants (*22*, *23*), digger wasps (*24*, *25*), bark beetles (*26*), darkling beetles (*27*, *28*), leaf rolling weevils (*29*), leaf beetles (*30*), and others, external symbiotic actinobacteria, proteobacteria, or ascomycetous fungi have been reported to produce potent antimicrobial substances, which protect their host insects against microbial parasites and pathogens (*31*–*33*). To discover whether this was the case for *M. gracilicorne*, we cocultured hindleg-derived *Simplicillium*, *Lecanicillium*, and *Akanthomyces* with several bacteria (*Escherichia coli*, *Pantoea dispersa*, and *Bacillus subtilis*), but we did not detect suppression or expression of antibiotic-related genes. RNA sequencing data of the hindlegs of mature adult females detected no expression from antibiotic-related genes, such as polyketide synthases (data S8). However, the possibility that hindleg-associated fungi were effective against hitherto unidentified microbial enemies of *M. gracilicorne* cannot be excluded.

Defense against wasp parasitism

Field-collected eggs of *M. gracilicorne* frequently suffer wasp parasitism. Whereas newly laid eggs of this stinkbug were white (Fig. 4A), becoming

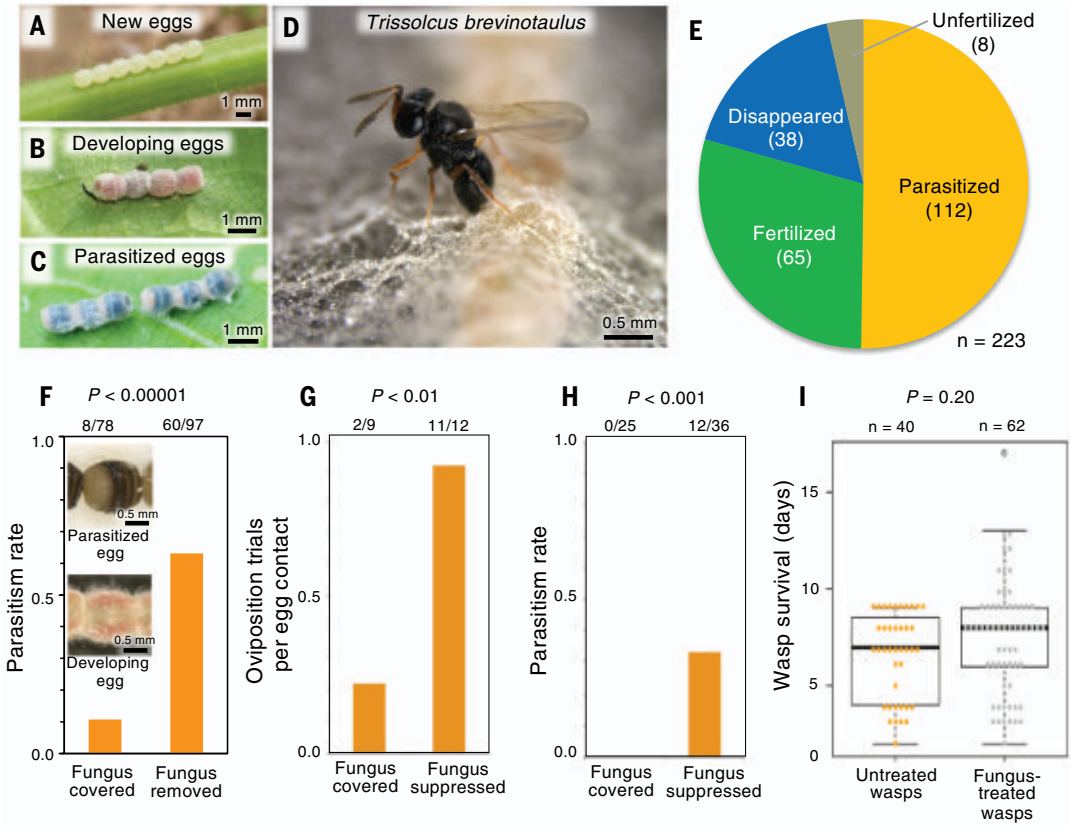


Fig. 4. Suppression of wasp parasitism by egg-covering fungal hyphae. (**A**) Newly laid eggs, white in color. (**B**) Developing fertilized eggs, reddish in color. (**C**) Wasp-parasitized eggs, black in color. (**D**) *T. brevinotaulus* ovipositing into an egg of *M. gracilicorne*. (**E**) Wasp parasitism rate in natural eggs of *M. gracilicorne*. In total, 50 newly laid egg masses consisting of 223 eggs were marked in the field and inspected 1 week later. (**F**) Parasitism success rates of *T. brevinotaulus* on fungus-covered eggs and fungus-removed eggs by brushing. The fungus removing procedure is shown in movie S3. (**G** and **H**) Oviposition trials per egg contact (G) and parasitism success rates (H) of *T. brevinotaulus* on fungus-covered eggs and fungus-suppressed eggs produced by ablation of maternal hindlegs. The fungus suppression procedure is shown in fig. S10, e to h. (**I**) Survival of *T. brevinotaulus* experimentally forced to contact with conidia of the hindleg-derived fungi. Significance levels are estimated by χ^2 test for (F) to (H) and generalized linear model test with a Poisson distribution for (I).

reddish within a week (Fig. 4B) and hatching within two weeks (fig. S2), some field-collected eggs were atypically black (Fig. 4C) because of wasp parasitism by *Trissolcus brevinotaulus* (*34*) (Fig. 4D). When we marked and monitored egg masses in the field, we found that more than half of the eggs were parasitized by the wasp (Fig. 4E). To evaluate the effect of the fungi on parasitism, we removed fungi from a subset of the egg masses by brushing of the hyphae and presented them to reproductive female wasps. In the experimental arena, the female wasps approached both the fungus-removed eggs and the fungus-covered eggs. Immediately after antennal drumming on the egg surface, the female wasps only oviposited on the cleaned eggs (movie S3). The female wasps were unable to oviposit at all on fully fungus-covered eggs (movie S4), but if eggs were sparsely coated with fungus, the female wasps were able to compress hyphae by antennal drumming and sometimes succeeded in oviposition (movie S5). The antennae of *T. brevinotaulus* are conspicuously thicker in females than in males, which may be not only for sensing host eggs but also for flattening fungal hyphae (fig. S10, a to d). Concordantly, fungus-removed eggs suffered significantly higher wasp parasitism than fungus-covered eggs (Fig. 4F). Removal of the hindlegs of female stinkbugs reduced the fungal covering on egg masses (fig. S10, e to h), which subsequently suffered more wasp parasitism than fungus-covered eggs (Fig. 4, G and H). It is notable that wasps still approached egg masses fully covered by fungal hyphae, despite showing intense self-grooming, which suggested that the hyphae were adherent. Fungal conidia were not able to infect the wasps (Fig. 4I). Thus, it appears that the defense offered to the eggs of *M. graciliforme* by fungal symbionts is physical rather than specifically chemical or pathogenic.

Recent invasion of a distinct parasitic wasp species

Our initial field and laboratory studies were conducted from 2014 to 2021, during which *T. brevinotaulus* was the only egg parasite observed. In 2024, we collected 275 adult insects of *M. gracilicorne* in the field and kept them in the laboratory to harvest egg masses, from which we recovered 212 normal fungus-covered egg masses and also generated 215 egg masses cleaned of fungus. Egg masses attached to paper were placed in the natural habitat of *M. gracilicorne*, and after 3 or 6 days field exposure, egg masses were recovered and the frequency of wasp parasitism was recorded with and without the presence of fungal hyphae (fig. S11, a to g, and data S9). About one-third of the eggs were preyed upon by snails (fig. S11, h and i, and data S9), but predation rates did not differ between fungus-covered eggs and the "cleaned" eggs (fig. S11j). Approximately 20% suffered wasp parasitism, in which higher parasitism rates were observed in the cleaned eggs than in the fungus-covered eggs (fig. S12a). Cleaned eggs exhibited lower *M. gracilicorne* hatching rates and higher mortality due to wasp parasitism (fig. S12, b and c). These results indicate that the fungal covering provides defense against wasp parasitism and contributes to host fitness under natural conditions. We compared parasitism rates within individual egg clusters, which also showed that the wasps attacked the fungus-removed eggs more frequently than the fungus-covered eggs (fig. S12d). Unexpectedly, smaller polyparasitic wasps, *Ooencyrtus* sp., emerged from the parasitized eggs more frequently than *T. brevinotaulus* (fig. S11, k to p, and data S9), which it appears to be replacing in Japan. Although *Ooencyrtus* sp. was effectively suppressed by the fungal symbionts, no suppression was detected for *T. brevinotaulus* in this field experiment (fig. S12, e and f), possibly because of limited sample sizes, environmental fluctuations in the field, or interactions with the invasive wasp species.

Discussion

On the basis of these results, we conclude that the fungi selectively cultured on the female's hindleg organ of *M. gracilicorne* are transferred to eggs to act as a physical defense against parasitic wasp attack. We have identified a specialized external cuticular structure bearing glandular mycangia for fostering specific fungal associates. It has been reported that the common cutworm moth *Spodoptera litura* lay eggs

densely covered with scale hairs, which prevent egg parasitism by the wasp *Trichogramma chilonis* (*35*). Such structural defense against natural enemies is also known from gall-forming insects, in which thick and voluminous gall tissues, often with conspicuous projections or hairs on the surface, prevent access and oviposition of parasitic wasps (*36*). In some gall midges, symbiotic fungi contribute to the hardness of external wall of galls, as well as serving as food for the fly larvae (*37*, *38*). Previous studies on defensive microbial symbioses in diverse insects have identified antibiotics (*23*, *25*–*28*), polyketide toxins (*39*, *40*), phage-encoded toxins (*41*), ribosome inactivating proteins (*42*), and other bioactive substances as microbial symbiont-derived chemical factors for defense (*31*–*33*). Although not detected in this study, the possibility remains that chemical defenses also play a part in the *M. gracilicorne* symbiosis.

Enlarged hindlegs are observed in males of several groups of heteropteran bugs and are used for courtship fighting and territorial behaviors (*43*, *44*). Meanwhile, female-specific ornamentation in hindlegs has only been seen so far among the Dinidoridae. We also inspected other dinidorid species, *Megymenum brevicorne*, *Coridius chinensis*, and *Cyclopelta parva* from Taiwan, and *Eumenotes pacao* from a subtropical island of Japan (data S10), all of which had the female-specific hindleg organ (fig. S13) and showed distinctive egg-smearing behavior during egg-laying (movies S6 to S9). These observations suggest that the hindleg organ and the egg-smearing behavior may have evolved early in the Dinidoridae (fig. S14). Fungal composition analysis of the hindleg organs of these species showed fungi growing on the hindleg organs of adult females of *M. brevicorne*, including *Simplicillium* and *Cordyceps* spp. Immature adult females of *M. brevicorne* exhibited diverse fungi at low densities on their hindleg organs. Diverse fungi were observed in *Co. chinensis*, *Cy. parva*, and *E. pacao* at low densities (fig. S15). These results suggest that the preferential culturing and association of the Cordycipitaceae fungi might have evolved in an ancestor of the genus *Megymenum* in the Dinidoridae (fig. S14). We speculate that the female-specific hindleg organ evolved for smearing secretions onto eggs for chemical defense or camouflage, but environmental fungi colonized the organ and exploited the secretion, and the insects co-opted the fungi for egg defense. The coexistence of vertical symbiont transmission, symbiont loss, and environmental symbiont reacquisition in the insect life cycle may be due to a trade-off between the benefits of defending eggs and the potential cost of infection. Future studies should focus on what molecular mechanisms underpin the development of the dinidorid symbiotic organ and how the specific fungal lineages are selected and cultivated by the insect.

REFERENCES AND NOTES

1. J. Strauß, R. Lakes-Harlan, in *Insect Hearing and Acoustic Communication*, B. Hedwig, Ed. (Springer, 2014), pp. 5–26.
2. P. S. S. Durai, *Orient. Insects* **21**, 163–360 (1987).
3. R. T. Schuh, J. A. Slater, *True Bugs of the World (Hemiptera: Heteroptera): Classification and Natural History* (Siri Scientific Press, ed. 2, 2020).
4. J. A. Lis, P. Lis, D. J. Ziaja, A. Kocorek, *Zootaxa* **3423**, 61–68 (2012).
5. J. A. Lis, *Pol. J. Entomol.* **72**, 297–304 (2003).
6. T. Kobayashi, S. Tachikawa, *Illustrated Eggs and Nymphs of Pentatomoids: Morphology and Ecology* (Yokendo, 2004) [In Japanese].
7. G.-H. Sung *et al.*, *Stud. Mycol.* **57**, 5–59 (2007).
8. R. M. Kepler *et al.*, *IMA Fungus* **8**, 335–353 (2017).
9. R. Zare, W. Gams, *Nova Hedwigia* **73**, 1–50 (2001).
10. Y. M. Zhou, J. R. Zhi, J. J. Qu, X. Zou, *Front. Microbiol.* **13**, 859886 (2022).
11. W. Chen *et al.*, *Front. Microbiol.* **13**, 950773 (2022).
12. M. Bright, S. Bulgheresi, *Nat. Rev. Microbiol.* **8**, 218–230 (2010).
13. H. Salem, L. Florez, N. Gerardo, M. Kaltenpoth, *Proc. Biol. Sci.* **282**, 20142957 (2015).
14. T. Hosokawa, T. Fukatsu, *Curr. Opin. Insect Sci.* **39**, 91–100 (2020).
15. C. Hamiaux, D. Stanley, D. R. Greenwood, E. N. Baker, R. D. Newcomb, *J. Biol. Chem.* **284**, 3496–3503 (2009).
16. G. M. Happ, C. M. Happ, S. J. Barras, *Tissue Cell* **3**, 295–308 (1971).
17. W. D. Stone, T. E. Nebeker, W. A. Monroe, J. A. MacGown, *Can. J. Zool.* **85**, 232–238 (2007).
18. C. R. Currie, M. Poulsen, J. Mendenhall, J. J. Boomsma, J. Billen, *Science* **311**, 81–83 (2006).
19. W. Goettler, M. Kaltenpoth, G. Herzner, E. Strohm, *Arthropod Struct. Dev.* **36**, 1–9 (2007).
20. J. Hulcr, L. L. Stelinski, *Annu. Rev. Entomol.* **62**, 285–303 (2017).
21. R. Koga *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2103957118 (2021).
22. C. R. Currie, J. A. Scott, R. C. Summerbell, D. Malloch, *Nature* **398**, 701–704 (1999).
23. D. C. Oh, M. Poulsen, C. R. Currie, J. Clardy, *Nat. Chem. Biol.* **5**, 391–393 (2009).
24. M. Kaltenpoth, W. Göttler, G. Herzner, E. Strohm, *Curr. Biol.* **15**, 475–479 (2005).
25. J. Kroiss *et al.*, *Nat. Chem. Biol.* **6**, 261–263 (2010).
26. J. J. Scott *et al.*, *Science* **322**, 63 (2008).
27. L. V. Flórez *et al.*, *Nat. Commun.* **8**, 15172 (2017).
28. L. V. Flórez *et al.*, *Nat. Commun.* **9**, 2478 (2018).
29. L. Wang *et al.*, *ISME J.* **9**, 1793–1801 (2015).
30. A. Berasategui *et al.*, *Curr. Biol.* **32**, 4114–4127.e6 (2022).
31. L. V. Flórez, P. H. W. Biedermann, T. Engl, M. Kaltenpoth, *Nat. Prod. Rep.* **32**, 904–936 (2015).
32. K. M. Oliver, A. H. Smith, J. A. Russell, *Funct. Ecol.* **28**, 341–355 (2014).
33. E. B. Van Arnam, C. R. Currie, J. Clardy, *Chem. Soc. Rev.* **47**, 1638–1651 (2018).
34. K. Matsuo, Y. Hirose, N. F. Johnson, *Jpn. J. Syst. Entomol.* **26**, 19–22 (2020).
35. I. Ohta, K. Miura, M. Kobayashi, *Appl. Entomol. Zool.* **29**, 608–609 (1994).
36. G. N. Stone, K. Schönrogge, *Trends Ecol. Evol.* **18**, 512–522 (2003).
37. A. E. Weis, *Ecology* **63**, 1602–1605 (1982).
38. J. O. Stireman 3rd, H. Devlin, T. G. Carr, P. Abbot, *Mol. Phylogenet. Evol.* **54**, 194–210 (2010).
39. J. Piel, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14002–14007 (2002).
40. A. Nakabachi *et al.*, *Curr. Biol.* **23**, 1478–1484 (2013).
41. K. M. Oliver, P. H. Degnan, M. S. Hunter, N. A. Moran, *Science* **325**, 992–994 (2009).
42. P. T. Hamilton, F. Peng, M. J. Boulanger, S. J. Perlman, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 350–355 (2016).
43. D. J. Emlen, H. F. Nijhout, *Annu. Rev. Entomol.* **45**, 661–708 (2000).
44. K. Okada, Y. Suzaki, Y. Okada, T. Miyatake, *Zool. Sci.* **28**, 659–663 (2011).
45. M. Moriyama, R. Koga, T. Hosokawa, N. Nikoh, H. Mukai, C.-H. Kuo, T. Fukatsu, T. Nishino, M. Tanahashi, S. Tachikawa, H. Chang, Defensive fungal symbiosis on insect hindlegs, Dryad (2025); https://doi.org/10.5061/dryad.66t1g1kcp.

ACKNOWLEDGMENTS

We thank R. Sugiyama, T. Taniguchi, W. Cai, C. Himuro, and Y.-H. Wu for collecting insect samples; K. Matsuo for identification of parasitic wasps; J. Araújo and H. Salem for reading and constructively commenting on the manuscript; and the core facilities of Academia Sinica (Taipei, Taiwan) for technical support, including the Cell Biology Core and the Genomic Technology Core (Institute of Plant and Microbial Biology) and the NGS High Throughput Genomics Core (Biodiversity Research Center). **Funding:** This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grants JP17H06388 (T.F., R.K.), JP24H02298 (T.F., M.M.), and JP19H02973 (M.M.); The Japan Science and Technology Agency (JST) ERATO Grant JPMJER1902 (T.F.); The JSPS Research Fellowship for Young Scientists (T.N); and Academia Sinica (C.-H.K.). **Author contributions:** Conceptualization: T.H., M.T., S.T., T.F.; Data curation: T.N., M.M., M.T., N.N.; Formal analysis: T.N., M.M., H.M., M.T., T.H., H.-Y.C., N.N., R.K.; Funding acquisition: T.F., R.K., M.M., T.N., C.-H.K.; Investigation: T.N., M.M., H.M., M.T., T.H.; Methodology: T.N., M.M., H.M., M.T., T.H., N.N., R.K.; Project administration: T.F., C.-H.K.; Resources: M.M., T.N., H.M., T.H., M.T., H.-Y.C., C.-H.K.; Supervision: T.F., M.M., R.K., C.-H.K.; Visualization: T.N., R.K.; Writing – original draft: T.F.; Writing – review & editing: T.F., T.N., M.M., H.M., M.T., C.-H.K. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Nucleotide sequence data obtained in this study were deposited in the DNA Data Bank of Japan (DDBJ) with the following accession numbers: fungal ITS sequence data, LC792825 to LC793396; Fungal ITS amplicon sequencing data, DRX509842 to DRX509916 and DRX637795 to DRX637852; RNA sequencing data, DRX509810 to DRX509841; insect mitochondrial genome data, LC670767 and PV423521 to PV423524. Raw data obtained in this study are available on Dryad (*45*).

License information: Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adp6699
Materials and Methods; Figs. S1–S15; References (*46*–*71*); Movies S1–S9; Data S1–S10; MDAR Reproducibility Checklist

Submitted 6 April 2024; resubmitted 30 April 2025; accepted 22 August 2025
10.1126/science.adp6699

PLANT ECOLOGY

Drought intensity and duration interact to magnify losses in primary productivity

Timothy Ohlert[1,2,*†], Melinda D. Smith[1,2,*†], Scott L. Collins[3], Alan K. Knapp[1,2], Jeffrey S. Dukes[4], Osvaldo Sala[5], Kate D. Wilkins[6], Seth M. Munson[7], Maggie I. Anderson[8], Meghan L. Avolio[9], Anping Chen[1,2], Meghan T. Hayden[10], Martin C. Holdrege[11], Ingrid J. Slette[12], Peter Wilfahrt[8], Claus Beier[13], Lauchlan H. Fraser[14], Anke Jentsch[15], Michael E. Loik[16], Yiqi Luo[17], Fernando T. Maestre[18], Richard P. Phillips[19], Sally A. Power[20], Laura Yahdjian[21],Qiang Yu[22,23,24]*, Angel Chen[25], Andrew J. Felton[26], Laureano A. Gherardi[27], Nicholas J. Lyon[25], Hamed Abdoli[28], Mehdi Abedi[28], Juan Alberti[29], Antonio I. Arroyo[30], Heidi Asbjornsen[31], Harald Auge[32,33], Seton Bachle[34], Michael Bahn[35], David C. Bartholomew[36], Amgaa Batbaatar[37,38], Taryn L. Bauerle[17], Karen H. Beard[11], Kai Behn[39], Ilka Beil[40], Lucio Biancari[18,21], Irmgard Blindow[40],Viviana Florencia Bondaruk[21], Elizabeth T. Borer[8], Edward W. Bork[38], Carlos Martin Bruschetti[29], Kerry M. Byrne[41], James F. Cahill Jr.[37], Dianela A. Calvo[42], Michele Carbognani[43], Cameron N. Carlyle[38], Karen Castillioni[8], Miguel Castillo-Garcia[30], Manjunatha H. Chandregowda[20], Scott X. Chang[44], Jeff Chieppa[20], Amber C. Churchill[45], Marcus Vinicius Cianciaruso[46], Amanda L Cordeiro[47], Sara A. O. Cousins[48], Daniela F. Cusack[47], Sven Dahlke[40], Pedro Daleo[29], Lee H. Dietterich[47,49], Maren Dubbert[50], Nico Eisenhauer[33,51], T'ai G. W. Forte[43,52], Flavia A. Funk[53], Darcy Galiano[54], Aaron C. Greenville[55], Liebao Han[22], Siri Vatsø Haugum[56], Yann Hautier[57], Andy Hector[58], Hugh A. L. Henry[59], Daniela Hoss[33,51], Forest Isbell[8], Samuel E. Jordan[7], Yuguang Ke[60], Eugene F. Kelly[61], Sally E. Koerner[62], Juergen Kreyling[40], György Kröel-Dulay[63], Alicia I. Kröpfl[64], Angelika Kübert[65,66], Andrew Kulmatiski[11], Eric G. Lamb[67], Klaus Steenberg Larsen[13], Steven Lee[68], Smriti Pehim Limbu[69], Anja Linstädter[70], Shirong Liu[71], Grisel Longo[72], Alejandro Loydi[53], Junwei Luan[73], F. Curtis Lubbe[74], Andrey V. Malyshev[40], Cameron D. McIntire[75], Daniel B. Metcalfe[36], Malesela Vincent Mokoka[76], Akira S. Mori[77], Edwin Mudongo[78], Gregory S. Newman[79], Uffe N. Nielsen[20], Raúl Ochoa-Hueso[80,81], Rory C. O'Connor[82], Romà Ogaya[83,84], Gastón R. Oñatibia[21], Ildikó Orbán[63,70], Brooke B. Osborne[85], Rafael Otfinowski[86], Meelis Pärtel[87], Jesús Pascual[29], Josep Peñuelas[83,84], Pablo L. Peri[88], David S. Pescador[89], Guadalupe Peter[42], Alessandro Petraglia[43], Catherine Picon-Cochard[90], Valério D. Pillar[91], Juan M. Piñeiro-Guerra[21,92], Laura Weber Ploughe[93], Robert M. Plowes[94], Cristy Portales-Reyes[95], Suzanne M. Prober[96], Yolanda Pueyo[30], Golsa Rahmati[28], Sasha C. Reed[97], Dana Aylén Rodríguez[53], William E. Rogers[98], Christiane Roscher[33,99], David W. Rowley[98,100], Ana M. Sánchez[101,102], Bráulio A. Santos[92], Michael P. Schellenberg[103], Michael Scherer-Lorenzen[104], Eric W. Seabloom[8], Ruonan Shen[105], Baoku Shi[106], Lara Souza[107], Andreas Stampfli[108], Rachel J. Standish[109], Marcelo Sternberg[110], Wei Sun[106], Marie Sünnemann[33,51], Michelle Tedder[111], Tyson J. Terry[15], Pål Thorvaldsen[112], Katja Tielbörger[113], Maud Tissink[35], Matthew A. Vadeboncoeur[31], Alejandro Valdecantos[114], Liesbeth van den Brink[113,115], Vigdis Vandvik[116], Liv Guri Velle[113], Svenja Wanke[15], Glenda M. Wardle[55], Cunzheng Wei[117], Christiane Werner[65], Georg Wiehl[118], Jennifer L. Williams[119], Amelia A. Wolf[120], Honghui Wu[24,121], Chong Xu[24,121], Xuechen Yang[122], Yadong Yang[24], Jenifer L. Yost[100], Alyssa L. Young[62], Ping Yue[123], Juan M. Zeberio[42], Michaela Zeiter[108,124], Haiyang Zhang[125], Juntao Zhu[126], Xiaoan Zuo[123]

As droughts become longer and more intense, impacts on terrestrial primary productivity are expected to increase progressively. Yet, some ecosystems appear to acclimate to multiyear drought, with constant or diminishing reductions in productivity as drought duration increases. We quantified the combined effects of drought duration and intensity on aboveground productivity in 74 grasslands and shrublands distributed globally. Ecosystem acclimation with multiyear drought was observed overall, except when droughts were extreme (i.e., ≤1-in-100-year likelihood of occurrence). Productivity losses after four consecutive years of extreme drought increased by ~2.5-fold compared with those of the first year. These results portend a foundational shift in ecosystem behavior if drought duration and intensity increase, from maintenance of reduced functioning over time to progressive and profound losses of productivity when droughts are extreme.

Drought, defined meteorologically as "a prolonged absence or marked deficiency of precipitation" (*1*), is a frequent and impactful disturbance in many terrestrial ecosystems globally. Although most droughts are short term and moderate in intensity (*2*), the most damaging and costly droughts from the perspective of ecological, societal, and economic impacts are both prolonged, unfolding over multiple years, and extreme with respect to long-term variation in climate conditions [e.g., (*3,4*)]. Although such drought events have historically occurred infrequently and, in some places, are absent from the recent historical record (*2*, *5*), there is evidence that longer-duration, intensified droughts are becoming more common (*6*, *7*) and will further increase in magnitude and frequency with global climate change (*5*, *8*, *9*). Yet, the impacts of multiyear, extreme droughts remain understudied, and past research is equivocal for how long-term droughts impact terrestrial ecosystems (*2*).

Theory predicts that as drought duration increases, the impacts of drought on ecosystem functioning (e.g., primary production) should accumulate or be magnified over time, resulting in more substantial losses in functioning, even for ecosystems that appear resistant to short-term drought (*2*, *10*). Several past studies report this expected cumulative pattern of response: a progressively more negative effect of drought on ecosystem functioning as duration increases (*11*, *12*). However, others find little evidence that increasing drought duration reduces functioning beyond that of a single-year drought [e.g., (*13*–*15*)]. Indeed, some research suggests that ecosystem function can "acclimate" or stabilize in response to multiyear drought [i.e., ecosystem acclimation; (*16*)], characterized by the impacts of drought remaining relatively constant or even diminishing over time (*11*, *16*–*19*). These variable responses to drought duration may result from differences in the magnitude (or intensity) of drought imposed. Indeed, drought duration and intensity are expected to interact in important ways (*2*, *10*). Droughts that are both prolonged and extreme are more likely to result in large impacts on ecosystem functioning (*10*, *20*, *21*). By contrast, short-term drought or prolonged moderate drought may result in lesser impacts on ecosystem functioning than extreme drought (*10*, *13*, *15*). Thus, to fully understand patterns of ecosystem response to drought duration, we need to also assess its interaction with drought intensity.

Our goals for this study were to 1) determine if prolonged drought results in a pattern of ecosystem response consistent with ecosystem acclimation (constant or lessening over time) vs. progressive losses (continuous decline over time), 2) quantify losses of ecosystem function attributable to each pattern, and 3) assess whether these

patterns of loss change with the magnitude of drought imposed. We achieved these goals with results from the International Drought Experiment (IDE), a multiyear global-scale study of drought effects on aboveground net primary productivity (hereafter referred to as "productivity"), a key measure of ecosystem functioning and a major component of the terrestrial carbon cycle (*22*).

The International Drought Experiment

The International Drought Experiment (IDE) is a coordinated drought experiment established in grassland and shrubland ecosystems across the globe [Fig. 1A and table S1; (*23*)]. These ecosystems cover ~40% of Earth's land surface, provide crucial ecosystem services [e.g., food, forage, fiber (*24*, *25*)], and their productivity is among the most responsive to

[1]Department of Biology, Colorado State University, Fort Collins, CO, USA. [2]Graduate Degree Program in Ecology, Colorado State University, CO, USA. [3]Department of Biology, University of New Mexico, Albuquerque, NM, USA. [4]Department of Global Ecology, Carnegie Institution for Science, Stanford, CA, USA. [5]School of Life Sciences, Arizona State University, Tempe, AZ, USA. [6]Denver Zoo Conservation Alliance, Denver, CO,, USA. [7]US Geological Survey, Southwest Biological Science Center, Flagstaff, AZ, USA. [8]Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, MN,, USA. [9]Morton K. Blaustein Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, MD, USA. [10]Department of Ecology & Evolutionary Biology, University of Colorado Boulder, CO, USA. [11]Department of Wildland Resources and Ecology Center, Utah State University, Logan, UT, USA. [12]Long-Term Ecological Research Network Office, National Center for Ecological Analysis and Synthesis, University of California Santa Barbara, Santa Barbara, CA, USA. [13]Department of Geosciences and Natural Resource Management, University of Copenhagen, Rolighedsvej 23, Frederiksberg C, Denmark. [14]Department of Natural Resource Science, Thompson Rivers University, Kamloops, BC, Canada. [15]Disturbance Ecology and Vegetation Dynamics, Bayreuth Center of Ecology and Environmental Research (BayCEER), Universitaetstr. 30, Bayreuth, University of Bayreuth, Germany. [16]Department of Environmental Studies, University of California, Santa Cruz, CA, USA. [17]School of Integrative Plant Science, Cornell University, Ithaca, NY, USA. [18]Environmental Sciences and Engineering, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia. [19]Department of Biology, Indiana University, 1001 East Third St., Bloomington, IN, USA. [20]Hawkesbury Institute for the Environment, Western Sydney University, Locked Bag 1797, Penrith, NSW, Australia. [21]Instituto de Investigaciones Fisiológicas y Ecológicas Vinculadas a la Agricultura (IFEVA), CONICET-Universidad de Buenos Aires, Facultad de Agronomía, Departamento de Recursos Naturales y Ambiente, Cátedra de Ecología, Buenos Aires, Argentina. [22]School of Grassland Science, Beijing Forestry University, Beijing, China. [23]State Key Laboratory of Efficient Production of Forest Resources, Beijing Forestry University, Beijing, China. [24]State Key Laboratory of Efficient Utilization of Arable Land in China, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing, China. [25]National Center for Ecological Analysis and Synthesis, University of California Santa Barbara, Santa Barbara, CA, USA. [26]Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, MT, USA. [27]Department of Environmental Sciences, Policy, and Management, University of California, Berkeley, CA, USA. [28]Department of Range Management, Faculty of Natural Resources and Marine Sciences, Tarbiat Modares University, Noor, Mazandaran Province, I. R. Iran. [29]Instituto de Investigaciones Marinas y Costeras (IIMyC), Universidad Nacional de Mar del Plata, CONICET, Argentina. [30]Pyrenean Institute of Ecology (IPE, CSIC), Montañana Avenue, Zaragoza, Spain. [31]Earth Systems Research Center, Department of Natural Resources and the Environment, University of New Hampshire, Durham, NH, USA. [32]Department of Community Ecology, Helmholtz-Center for Environmental Research - UFZ, Theodor-Lieser-Strasse 4, Halle, Germany. [33]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Puschstrasse 4, Leipzig, Germany. [34]LI-COR Environmental, Lincoln, NE, USA. [35]Department of Ecology, University of Innsbruck, Sternwartestr. 15, Innsbruck, Austria. [36]Department of Ecology and Environmental Science, Umeå University, Linnaeus väg 6, Umeå, Sweden. [37]Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada. [38]Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, Alberta, Canada. [39]Institute of Crop Science and Resource Conservation, University of Bonn, Bonn, Germany. [40]Institute for Botany and Landscape Ecology, University of Greifswald, Greifswald, Germany. [41]Department of Environmental Science and Management, California State Polytechnic University, Humboldt, Arcata, CA, USA. [42]Universidad Nacional de Río Negro, Sede Atlántica, CEANPa, CONICET, Viedma, Río Negro, Argentina. [43]Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parco Area delle Scienze 11/A, Parma, Italy. [44]Department of Renewable Resources, University of Alberta, Edmonton, Alberta, Canada. [45]Environmental Studies Program, Binghamton University, Binghamton, NY, USA. [46]Department of Ecology, Universidade Federal de Goiás, GO, Brazil. [47]Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, CO, USA. [48]Department of Physcial Geography, Stockholm University, Stockholm, Sweden. [49]Department of Biology, Haverford College, Haverford, PA, USA. [50]Leibniz Center for Agricultural and Landscape Research (ZALF), Müncheberg, Germany. [51]Institute of Biology, Leipzig University, Puschstrasse 4, Leipzig, Germany. [52]Department of Food and Drug, University of Parma, Parco Area delle Scienze 11/A, Parma, Italy. [53]Centro de Recursos Naturales Renovables de la Zona Semiárida (CERZOS) – CONICET, Departamento de Biología, Bioquímica y Farmacia, Universidad Nacional del Sur, Bahía Blanca, Argentina. [54]Asociación Civil Sin Fines De Lucro Para La Biodiversidad, Investigación Y Desarrollo Ambiental En Ecosistemas Tropicales (ABIDA), Urbanización Ucchullo Grande, Urbanización Ucchullo Grande, Avenida Argentina F-9, Cusco, Perú. [55]Sydney Institute of Agriculture, ARC Training Centre in Data Analytics for Resources and Environments (DARE), School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia. [56]The Heathland Centre at Lygra, Lygra, Norway. [57]Ecology and Biodiversity Group, Department of Biology, Utrecht University, Padualaan 8, Utrecht, The Netherlands. [58]Department of Biology (and Leverhulme Centre for Nature Recovery), University of Oxford, Oxford, UK. [59]Department of Biology, University of Western Ontario, London, ON, Canada. [60]Hulunber Grassland Ecosystem National Observation and Research Station, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing, China. [61]Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO, USA. [62]Department of Biology, University of North Carolina Greensboro, 321 McIver Street, 312 Eberhart Building, Greensboro, NC, USA. [63]HUN-REN Centre for Ecological Research, Institute of Ecology and Botany, 2-4 Alkotmány u., Vácrátót, Hungary. [64]Departamento de Gestión Agropecuaria, Universidad Nacional del Comahue, Cemtro Regional Universitario Zona Atlántica, Viedma, Argentina. [65]Ecosystem Physiology, University of Freiburg, Freiburg, Germany. [66]Institute for Atmospheric and Earth System Research/Physics, Faculty of Science, University of Helsinki, Helsinki, Finland. [67]Department of Plant Sciences, University of Saskatchewan, Saskatoon, Saskatchewan, Canada. [68]US Geological Survey, Western Ecological Research Center, Wawona, CA, USA. [69]Department of Environmental Studies, Dartmouth College, Hanover, NH, USA. [70]Department of Biodiversity Research/ Systematic Botany, University of Potsdam, Maulbeerallee 1, Potsdam, Germany. [71]Key Laboratory of Forest Ecology and Environment of National Forestry and Grassland Administration, Ecology and Nature Conservation Institute, Chinese Academy of Forestry, Beijing, China. [72]Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Universidade Federal da Paraíba, João Pessoa, Paraíba, Brasil. [73]Institute of Resources and Environment, Key Laboratory of Bamboo and Rattan Science and Technology of State Forestry and Grassland Administration, International Centre for Bamboo and Rattan, Beijing, China. [74]Department of Experimental and Functional Morphology, Institute of Botany, Czech Academy of Sciences, Třeboň, Czech Republic. [75]USDA Forest Service, Forest Health Protection, Durham, NH, USA. [76]University of Pretoria, Department of Geography, Geoinformatics and Meteo., South Africa. [77]Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan. [78]Communities Living Among Wildlife Sustainably (CLAWS) Botswana, PO Box121, Seronga, Botswana. [79]School of Biological Sciences, University of Oklahoma, Norman, OK, USA. [80]Department of Biology, IVAGRO, University of Cádiz, Campus de Excelencia Internacional Agroalimentario (ceiA3), Campus del Rio San Pedro, Puerto Real, Cádiz, Spain. [81]Department of Terrestrial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), P.O. Box50, Wageningen, The Netherlands. [82]Range and Meadow Forage Management Research Unit, Eastern Oregon Agricultural Research Center, USDA-Agricultural Research Service, Burns, OR, USA. [83]Global Ecology Unit CREAF-CSIC-UAB, CSIC, Bellaterra, Catalonia, Spain. [84]CREAF, Cerdanyola del Vallès, Barcelona, Catalonia, Spain. [85]Department of Environment and Society, Utah State University, Moab, UT, USA. [86]Department of Biology, The University of Winnipeg, 515 Portage Ave, Winnipeg, Manitoba, Canada. [87]Institute of Ecology and Earth Sciences, University of Tartu, J. Liivi 2, Tartu, Estonia. [88]Instituto Nacional de Tecnología Agropecuaria (INTA)-Universidad Nacional de la Patagonia Austral (UNPA)-CONICET, Río Gallegos, Santa Cruz, Argentina. [89]Departamento de Farmacología, Farmacognosia y Botánica, Facultad de Farmacia, Universidad Complutense de Madrid, Madrid, Spain. [90]Université Clermont Auvergne, INRAE, VetAgro Sup, UREP, Clermont-Ferrand, France. [91]Department of Ecology, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil. [92]Departamento de Sistemática e Ecologia, Universidade Federal da Paraíba, João Pessoa, Paraíba, Brasil. [93]National Park Service, Inventory and Monitoring Division, Flagstaff, AZ, USA. [94]Brackenridge Field Laboratory, University of Texas at Austin, Austin, TX, USA. [95]Department of Biology, Saint Louis University, St. Louis, MO, USA. [96]CSIRO Environment, PO Box1700, Canberra ACT, Australia. [97]U.S. Geological Survey, Southwest Biological Science Center, Moab, UT, USA. [98]Department of Ecology and Conservation Biology, Texas A&M University, College Station, TX, USA. [99]Department of Physiological Diversity, Helmholtz-Cente for Environmental Research - UFZ, Permoserstrasse 15, Leipzig, Germany. [100]USDA-ARS Grassland, Soil and Water Research Laboratory, Temple, TX, USA. [101]Instituto de Investigación en Cambio Global (IICG-URJC), Universidad Rey Juan Carlos, 28933 Móstoles, Madrid, Spain. [102]Departamento de Biología y Geología, Física y Química Inorgánica, URJC, 28933 Móstoles, Madrid, Spain. [103]Swift Current Research and Development Centre, Agriculture and Agri-Food Canada, 1 Airport Rd, Swift Current, Saskatchewan, Canada. [104]Faculty of Biology, University of Freiburg, Schänzlestraße 1, Freiburg, Germany. [105]Key Laboratory of Ecosystem Network Observation and Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China. [106]Institute of Grassland Science, Key Laboratory of Vegetation Ecology of the Ministry of Education, Jilin Songnen Grassland Ecosystem National Observation and Research Station, Northeast Normal University, Changchun, China. [107]Oklahoma Biological Survey & Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK, USA. [108]School of Agricultural, Forest and Food Sciences, Bern University of Applied Sciences, Zollikofen, Switzerland. [109]School of Environmental and Conservation Sciences, Murdoch University, 90 South Street, Murdoch, Western Australia. [110]School of Plant Sciences & Food Security, Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel. [111]School of Life Sciences, University of Kwazulu-Natal, Pietermaritzburg, South Africa. [112]Norwegian Institute of Bioeconomy Research, Tjøtta, Norway. [113]Plant Ecology Group, Department of Biology, University of Tübingen, Tübingen, Germany. [114]Department of Ecology and IMEM, University of Alicante, Alicante, Spain. [115]ECOBIOSIS, Department of Botany, University of Concepcion, Concepcion, Chile. [116]Department of Biological Sciences and Bjerknes Centre for Climate Research, University of Bergen, Norway. [117]State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing, China. [118]CSIRO Environment, 147 Underwood Avenue, Floreat, WA, Australia. [119]Department of Geography and Biodiversity Research Centre, University of British Columbia, Vancouver, BC, Canada. [120]Department of Integrative Biology, University of Texas at Austin, Austin, TX, USA. [121]Key Laboratory of Arable Land Quality Monitoring and Evaluation, Ministry of Agriculture and Rural Affairs, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing, China. [122]State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, China. [123]State Key Laboratory of Ecological Safety and Sustainable Development in Arid Lands, Urat Desert-grassland Research Station, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou, China. [124]Institute of Plant Sciences, University of Bern, Bern, Switzerland. [125]School of Life Sciences, Institute of Life Science and Green Development, Hebei University, Baoding, China. [126]Lhasa Plateau Ecosystem Research Station, Key Laboratory of Ecosystem Network Observation and Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China. ***Corresponding author. Email: melinda.smith@colostate.edu (M.D.S.); timothy.ohlert@colostate.edu (T.O.); yuq@bjfu.edu.cn (Q.Y.)** †These authors contributed equally to this work.

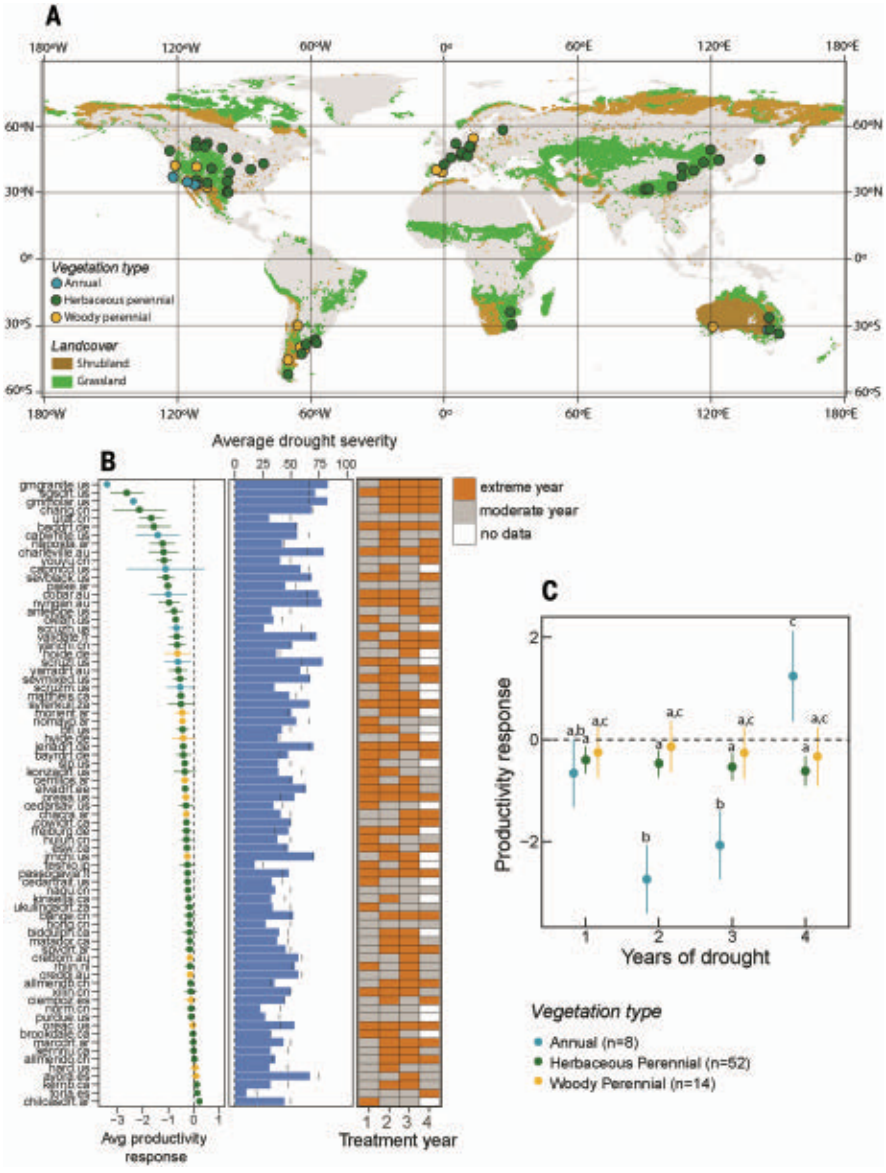


Fig. 1. Overview of the IDE: geographic locations, drought treatments, and effects on aboveground productivity. (**A**) Locations of the 74 IDE sites included in this study and their distribution across six continents (site names are given in table S1). Background shading denotes Moderated Resolution Imaging Spectroradiometer–derived landcover types (*50*), and the colors of the points denote the vegetation type of each site: annual, herbaceous perennial, or woody perennial (*23*). (**B**) (Left) IDE sites ordered by the average productivity response to drought over the 3- to 4-year duration of the experiment. Error bars represent the standard error for each site. (Middle) The average drought severity [defined as (MAP − precipitation received by drought treatment plots)/MAP; (*23*)] experienced over the duration of the experiment (blue bars). The expected average drought severity for the target 1-in-100–year drought treatment is indicated by the vertical black line. Overall, 53% of sites experienced an average precipitation reduction equivalent to the level expected with the target 1-in-100–year extreme drought treatment over the duration of the experiment. (Right) The temporal sequence of extreme (orange) versus moderate (gray) drought years imposed at each site. Note that 21 sites imposed only 3 years of drought treatment, and, therefore, the designation for the fourth year of treatment is left empty (white). (**C**) Average productivity response to drought (moderate and extreme combined) over time for three vegetation types. Productivity response was calculated as the natural log of the ratio of productivity during drought to long-term mean productivity (*23*). For example, a productivity response of –1 equates to a change in productivity due to drought of about 63% of the long-term mean. Error bars represent standard error, and letters denote statistical differences among groups based on a linear mixed effect model and post hoc comparison (table S3).

precipitation variability [e.g., (*26*)]. IDE sites were established on six continents and span broad precipitation, temperature, and environmental gradients [Fig. 1A and table S1; (*27*, *28*)]. All IDE sites utilize a common experimental approach: passive rainfall manipulation shelters (*29*) that simulated year-round drought [365 days; (*27*)] for up to 4 years. This allows for drought-duration impacts to be assessed in a cost-effective manner while still representing key characteristics of natural drought events [i.e., smaller and fewer rainfall events accompanied by longer periods between rainfall events; (*29*)]. At the time of this analysis, there were 74 grassland and shrubland IDE sites that had imposed 3 ($n = 21$) or 4 ($n = 53$) years of drought (Fig. 1B).

In addition to drought duration, IDE was designed to capture another way in which drought events are changing: increased intensity (or magnitude). To accomplish this, we selected a statistically extreme target level of drought intensity tailored for each IDE site: a 1-in-100–year drought based on long-term annual precipitation amounts available from site-level historical records (Fig. 1B) (*23*). By choosing this target level, our intent was to impose a scenario of extreme drought that is currently predicted to become more common with climate change in the near future, yet not so extreme as to be unrealistic [e.g., a 1-in-100–year drought will become more common well before a 1-in-1000–year drought does; (*30*)]. Thus, the goal with IDE was to apply drought treatments that (i) were historically and statistically rare for most if not all sites included in our study but also (ii) are forecast to become more common with climate change (*31*).

The IDE passive rainfall manipulation shelters rely on ambient precipitation to achieve drought (*29*). However, because ambient precipitation varied each year of the study, the target 1-in-100–year drought treatment was realized only when ambient rainfall was less than or equal to mean annual precipitation (MAP) for a site (*23*). When this criterion was met, we categorized the drought treatment as "extreme" [following (*27*)]. By contrast, when ambient annual rainfall was greater than MAP for a site, the target 1-in-100–year drought was not met, but drought was still imposed. For this scenario, we categorized the drought treatment as "moderate." The extreme and moderate categories of drought intensity align with those used in well-recognized drought classification systems, such as the US Drought Monitor (*23*). We also quantified the IDE drought treatments as a continuous variable using a common and comparable drought severity metric (*32*), calculated as the relative reduction in rainfall in the drought treatment from MAP (*23*). Average drought severity was substantially greater (~60%) for the extreme versus moderate drought intensity categories (fig. S1). An additional feature of the IDE design is that, in any given year, approximately half of the sites experienced extreme, 1-in-100–year drought, and after multiple years, sites experienced different combinations of moderate and extreme drought years (Fig. 1B). This allowed us to contrast distinct sequences of moderate and extreme drought impacts over multiple years.

Variability in drought response over time

Previously, we showed that average productivity reductions were ~60% greater when a single-year drought was extreme versus not extreme; however, variability among IDE sites in their response to short-term drought was notably high, ranging from complete resistance (i.e., no reduction in productivity) to large declines in productivity (*27*). Much of this variability in response was related to variation in drought severity, with productivity decreasing, as expected, with increasing drought severity (*27*). As drought duration was increased from 1 to 4 years in this study, we expected that variation in productivity responses among sites would decrease. However, average productivity responses to multiple (3 to 4) years of drought remained notably variable, ranging from little response to as much as a 97% decline in productivity (Fig. 1B).

We examined a broad set of biotic and abiotic variables previously hypothesized to explain variation in drought response (*23*), including differences in plant species richness, abundance of key growth forms (i.e., graminoids), soil texture, MAP, mean annual temperature, mean aridity index (AI), interannual precipitation variability, precipitation seasonality, and previous and current year drought severity (figs. S2 to S5 and table S2). We found that, as drought extended over multiple years, previous years' drought severity (years 2 and 3), MAP (years 2 to 4), mean AI (year 3), interannual variation (year 4) and seasonality in precipitation (year 3), and plant species richness (years 1 and 4) were major predictors of variation in drought response (figs. S2 to S5). Thus, as found in other studies (*27*, *33*), drier and less biodiverse sites, as well as those with more variable or more seasonal precipitation, experienced greater losses in productivity with drought. However, drought severity was the best and most consistent predictor of variation in drought response, as observed with single-year droughts (*27*).

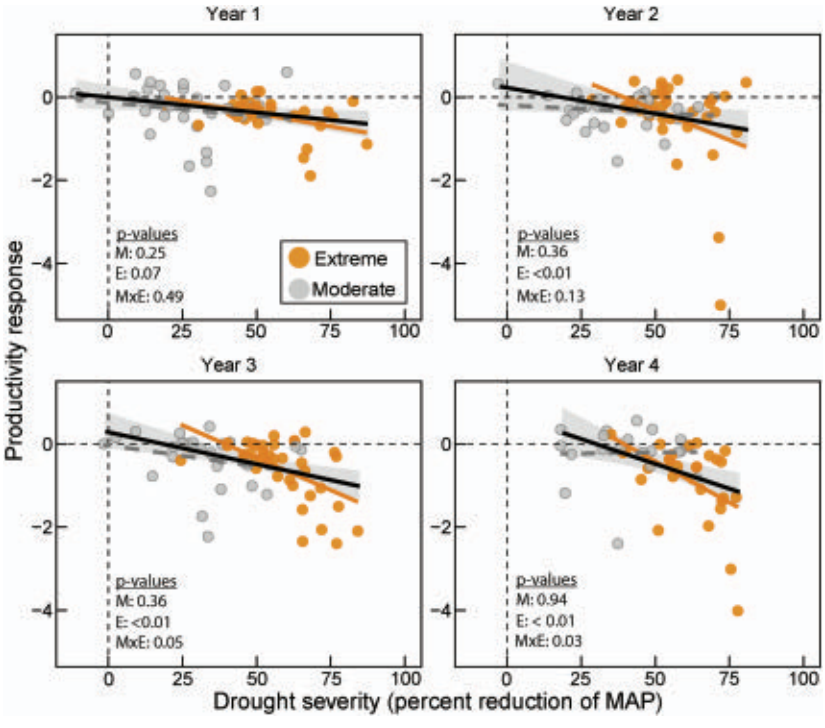


Fig. 2. Relationships between ecosystem productivity response to drought and drought severity across all sites (black line) and moderate (gray dots) versus extreme (orange dots) drought intensities for each of the 4 years of the experiment. Productivity response was calculated as the natural log of the ratio of productivity during drought to long-term mean productivity (*23*). Drought severity was calculated as: (MAP − precipitation received by drought treatment plots)/MAP (*23*). The regression across all sites was significant for all years (table S4). *P* values for moderate and extreme regressions are shown in the bottom left corner of each panel (table S5): M, moderate regression; E, extreme regression; MxE, the interaction between moderate and extreme regressions (i.e. whether the slopes differ from each other).

Pattern of productivity loss with multiyear drought

Despite variation in drought response among sites, we expected that a pattern of progressive (or cumulative) losses of productivity would emerge at most sites as drought continued over multiple years. After a significant decline in productivity in the first year of drought (29%), when averaged across all sites and drought intensity categories, productivity did not continue to decrease over time (Fig. 1C and table S3). Instead, ecosystem acclimation was generally observed. Notably, annual grasslands responded distinctly from perennial grasslands and shrublands, exhibiting a much larger initial response, but with the response lessening over time (table S3). Previous studies in annual-dominated systems have also found similar responses as well as strong drought resistance (*34*, *35*). Unfortunately, given the small number of annual-dominated IDE sites ($n = 8$) and their limited geographic coverage (seven were in the southwestern United States, and six experienced above-average precipitation in year 4), it is difficult to draw substantive conclusions about the nature of drought-duration effects based on these annual ecosystems. As such, we focused all subsequent analyses on the more widely represented perennial-dominated grassland and shrubland sites.

Interaction of drought duration and severity

The above analysis of drought duration effects does not consider intensity (extreme versus moderate) of the drought imposed. However, we expected that losses in productivity under extreme drought would be magnified over time and most pronounced when drought intensity was consistently extreme over multiple years. We tested this prediction in four ways. First, we examined relationships between productivity responses and drought severity for each year of the drought using multimodel comparisons that also included the previous year's drought severity to account for potential carry-over effects of the severity of drought from one year to the next (*23*). Consistent with simple linear regression analyses (figs. S2 to S5 and table S2), the current year's drought severity was the best predictor of variation in productivity response, regardless of ecosystem type or the previous year's drought severity (Fig. 2 and table S4). However, this analysis does not consider whether drought intensity was extreme or not. Therefore, we tested whether the slope of the relationship between productivity responses and drought severity would change depending on whether drought intensity was extreme versus moderate. This analysis allowed us to consider the magnitude of the drought treatment as both a continuous (i.e., severity) and categorical variable when describing the productivity response to drought over time. We found that, by year 3, the relationship between drought severity and productivity responses differed significantly between moderate versus extreme droughts, with the difference in these relationships most pronounced in year 4 (Fig. 2 and table S5). In other words, the slope of the relationship between drought severity and productivity loss became more negative over time when the intensity of the drought treatment was extreme, whereas the slope of the relationship for moderate intensity droughts did not change significantly over time. Third, we assessed the impact of extreme versus moderate drought intensity in any given year during the 4-year period of precipitation reductions, regardless of the previous year's drought severity. We found that average productivity losses significantly increased over time when drought was extreme, whereas the effects of moderate intensity droughts on productivity were independent of the year in which they occurred (Fig. 3A and table S6). Lastly, we quantified productivity responses of sites that had experienced only extreme drought for 1 to 4 years. For the subset of sites

with such consecutive extreme drought years, the strongest duration effects were revealed (Fig. 3B and table S7), with a ~2.5-fold greater loss of productivity as duration increased from 1 to 4 years (a 29 versus 77% reduction, respectively). Collectively, these results support predictions that droughts of extreme intensity cause greater impacts on ecosystem functioning than moderate droughts of similar duration (*10*). However, most notably, we show that increasing drought duration concurrent with consistently extreme drought results in progressive losses in ecosystem functioning that are more profound than previously reported (*11*, *12*).

Conclusions

Our results help to reconcile contrasting patterns of drought duration responses reported previously. IDE results show that, after an initial loss of function in year one, ecosystems subjected to multiple years of moderate (or less severe) drought are likely to maintain this level of limited functioning (i.e., exhibit ecosystem acclimation). By contrast, an increase in severity to historically extreme levels will result in a pattern of cumulative loss of function over time. There are several mechanisms that may result in patterns of ecosystem acclimation versus cumulative effects of drought (*10*, *20*, *21*), including demographic and community shifts resulting from mortality or establishment failure (leading to loss in function) as well as plastic or adaptive responses to drought over time (leading to mitigation of loss over time). Although the IDE was not designed to rigorously test such mechanisms, available data from 49 sites on species gains and losses as well as changes in species richness suggest that demographic and community shifts likely occurred (fig. S6A and tables S8 and S9), and over time, greater species losses were significantly related to increased losses in productivity with drought (fig. S6B and table S10). Although additional research will be required to test mechanisms that may determine acclimation versus cumulative responses to drought, such mechanistic understanding is crucial in a future where extreme droughts become the norm.

The lack of duration effect with moderate drought intensity is not entirely surprising, given that many grassland and shrubland ecosystems occur in a broad range of semiarid to arid climates, as did a majority of IDE sites (table S2). The ability of these water-limited systems to rapidly respond to short-term fluctuations in precipitation (*22*, *36*, *37*) but also maintain functioning for more extended dry periods is consistent with the long-term stability of these ecosystems (*38*). Indeed, it is also worth highlighting that a subset of sites was resistant to multiple years of drought, regardless of severity. It may be that these ecosystems are less water limited (table S2) and therefore less impacted by drought, as has been observed for mesic grasslands [e.g., (*14*,*16*)]. However, it should also be noted that drought experiments may underestimate drought effects (*39*), and although passive rainout shelters alter precipitation inputs and soil moisture in ways that accurately simulate changes in rainfall during natural droughts (*28*), they do not reproduce ancillary drought attributes, such as higher temperatures and vapor pressure deficits that typically accompany drought events (*40*–*42*). Although direct temperature effects are not particularly strong in grasslands (*43*, *44*), an increase in vapor pressure deficits during drought has the potential to reduce photosynthesis and productivity (*45*, *46*), and the lack of a temperature manipulation in this study could partially explain why some IDE sites were unresponsive to drought over time.

The discovery that the resistance to drought duration of grasslands and shrublands rapidly eroded with prolonged drought of extreme intensity portends an uncertain future for these ecosystems, threatening their long-term stability and the ecosystem goods and services they provide. Particularly alarming were the 160%-greater (or 2.5-fold–greater) reductions in productivity observed when extreme drought years occur consecutively. Extreme, consecutive drought years, including megadroughts (*8*), are expected to increase in the future with climate change (*8*,*31*). Although concerns about ecosystem stability in the face of ongoing increases in both drought magnitude and duration have been voiced for decades (*47*, *48*), our results provide experimental evidence in support of recent observations (*5*) that the functioning of these globally important ecosystems are at risk from longer and more intense droughts.

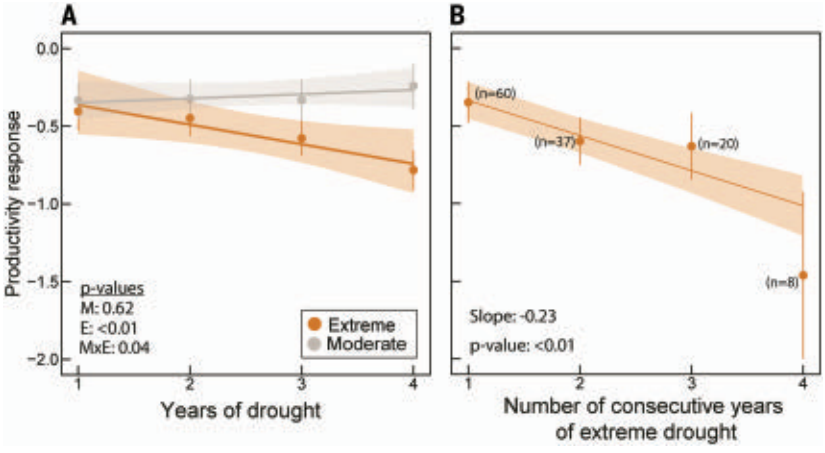


Fig. 3. Effects of drought duration (1 to 4 years) on productivity responses for moderate versus extreme drought intensities. (**A**) Despite an initial loss of productivity in the first year, drought duration had no effect on productivity responses when drought was moderate (gray line), irrespective of whether drought was of extreme or moderate intensity in previous years. By contrast, drought duration increased productivity losses when drought was extreme (orange line), irrespective of whether previous years were extreme or moderate. Thus, an extreme drought in year four reduced productivity more than an extreme drought in year one. *P* values for moderate and extreme regressions are shown in the bottom left corner of each panel (table S5): M, moderate regression; E, extreme regression; MxE, the interaction between moderate and extreme regressions (i.e. whether the slopes differ from each other). (**B**) Drought duration had much greater cumulative impacts on productivity if the years were consecutively extreme. The slope of the relationship between time and productivity response for consecutive extreme drought was twofold greater than in (A). Productivity responses to consecutive moderate droughts could not be assessed owing to an insufficient number of sites experiencing 3 ($n = 7$) and 4 ($n = 3$; see Fig. 1B) consecutive years of moderate drought. In both (A) and (B), productivity response was calculated as the natural log of the ratio of productivity during drought to long-term mean productivity (*23*), error bars represent standard error, and shading represents the 95% confidence interval of the regression. Summary statistics for the linear mixed effects model are shown in the bottom left corner: slope and *P* value of the regression (tables S6 and S7).

REFERENCES AND NOTES

1. World Meteorological Organization, *International Meteorological Vocabulary, WMO-No. 182* (WMO, ed. 2, 1992), p. 784.
2. A. J. Felton, A. K. Knapp, M. D. Smith, *Glob. Chang. Biol.* **27**, 1127–1140 (2021).
3. J. E. Weaver, *North American Prairie* (Johnson Publishing Company, 1954).
4. A. I. Van Dijk *et al.*, *Water Resour. Res.* **49**, 1040–1057 (2013).
5. L. Chen *et al.*, *Science* **387**, 278–284 (2025).
6. K. E. King *et al.*, *Sci. Adv.* **10**, eadj4289 (2024).
7. J. F. Silva, R. M. da Silva, C. A. Santos, A. M. Silva, P. C. Vianna, *Nat. Hazards* **108**, 1347–1371 (2021).
8. B. I. Cook *et al.*, *Nat. Rev. Earth Environ.* **3**, 741–757 (2022).
9. S. M. Vicente-Serrano *et al.*, *Philos. Trans. A Math. Phys. Eng. Sci.* **380**, 20210285 (2022).
10. M. D. Smith, *J. Ecol.* **99**, 656–663 (2011).
11. J. Gao, L. Zhang, Z. Tang, S. Wu, *J. Ecol.* **107**, 2519–2531 (2019).
12. C. Xu *et al.*, *Ecol. Indic.* **131**, 108139 (2021).
13. A. Jentsch *et al.*, *J. Ecol.* **99**, 689–702 (2011).
14. C. J. W. Carroll *et al.*, *Oecologia* **197**, 1017–1026 (2021).
15. A. Batbaatar, C. N. Carlyle, E. W. Bork, S. X. Chang, J. F. Cahill Jr., *J. Ecol.* **110**, 197–209 (2022).
16. Q. Yu *et al.*, *Nature* **639**, 114–118 (2025).
17. S. Backhaus *et al.*, *Ecosystems* **17**, 1068–1081 (2014).
18. M. Estiarte *et al.*, *Glob. Chang. Biol.* **22**, 2570–2581 (2016).
19. J. Wang *et al.*, *Glob. Ecol. Biogeogr.* **30**, 1909–1921 (2021).
20. V. P. Gutschick, H. BassiriRad, *New Phytol.* **160**, 21–42 (2003).
21. L. W. Ploughe *et al.*, *New Phytol.* **222**, 52–69 (2019).
22. A. K. Knapp, M. D. Smith, *Science* **291**, 481–484 (2001).

23. See materials and methods in the supplementary materials for details.
24. Y. Zhao, Z. Liu, J. Wu, *Landsc. Ecol.* **35**, 793–814 (2020).
25. H. Liu, L. Hou, N. Kang, Z. Nan, J. Huang, *Grassl. Res.* **1**, 63–74 (2022).
26. T. E. Huxman *et al.*, *Nature* **429**, 651–654 (2004).
27. M. D. Smith *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2309881120 (2024).
28. L. Yahdjian *et al.*, *Bioscience* **71**, 918–927 (2021).
29. A. K. Knapp *et al.*, *Glob. Chang. Biol.* **23**, 1774–1782 (2017).
30. L. Korell, H. Auge, J. M. Chase, S. Harpole, T. M. Knight, *Glob. Chang. Biol.* **26**, 325–327 (2020).
31. IPCC, *Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2022).
32. M. D. Smith, K. R. Wilcox, S. A. Power, D. T. Tissue, A. K. Knapp, *J. Veg. Sci.* **28**, 235–237 (2017).
33. V. F. Bondaruk *et al.*, *Nat. Ecol. Evol.* **9**, 937–946 (2025).
34. J. C. Ruppert *et al.*, *Glob. Chang. Biol.* **21**, 1258–1270 (2015).
35. K. Tielbörger *et al.*, *Nat. Commun.* **5**, 5102 (2014).
36. S. M. Vicente-Serrano *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 52–57 (2013).
37. L. A. Gherardi, O. E. Sala, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 12735–12740 (2015).
38. J. S. Clark *et al.*, *Ecology* **83**, 595–601 (2002).
39. G. Kröel-Dulay *et al.*, *Nat. Ecol. Evol.* **6**, 540–545 (2022).
40. A. K. Knapp *et al.*, *Funct. Ecol.* **38**, 76–97 (2024).
41. A. J. Wright, S. L. Collins, *Bioscience* **74**, 65–71 (2024).
42. J. Kreyling *et al.*, *Ecosystems* **20**, 301–315 (2017).
43. D. L. Hoover, A. K. Knapp, M. D. Smith, *Ecology* **95**, 2646–2656 (2014).
44. W. Mowll *et al.*, *Oecologia* **177**, 959–969 (2015).
45. A. G. Konings, A. P. Williams, P. Gentine, *Nat. Geosci.* **10**, 284–288 (2017).
46. C. Grossiord *et al.*, *New Phytol.* **226**, 1550–1566 (2020).
47. D. R. Easterling *et al.*, *Science* **289**, 2068–2074 (2000).
48. H. Yang *et al.*, *Glob. Chang. Biol.* **29**, 2351–2362 (2023).
49. T. Ohlert, M. D. Smith, Drought intensity and duration interact to magnify losses in primary productivity. Available on *DRYAD:* doi.org/10.5061/dryad.np5hqc05h
50. M. Friedl, D. Sulla-Menashe, MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 0.05Deg CMG V061, NASA EOSDros. Inf. Serv. Land Processes Distributed Active Archive Center (2022); https://doi.org/10.5067/MODIS/MCD12C1.061.

ACKNOWLEDGMENTS

We are grateful to the many IDE collaborators who established and maintained the IDE experiments, collected field data, and shared their data with the IDE community. **Funding:** Research support was provided by the following: National Science Foundation (NSF) Research Coordination Network grant to M.D.S. (DEB-1354732); US Department of Agriculture's National Institute of Food and Agriculture (USDA-NIFA) Postdoctoral Fellowship grant to K.D.W. (2020-67034-31898); USDA-NIFA Conference grant to M.D.S. (2020-67019-31757); US Geological Survey John Wesley Powell Center for Analysis and Synthesis grant to K.D.W. and M.D.S.; US Geological Survey grant to M.D.S. (G21AC10266-00); Global Drought Synthesis Group grant to K.D.W. and M.D.S., funded by the NSF Long-term Ecological Research Network Office (LNO) and the National Center for Ecological Analysis and Synthesis, University of California-Santa Barbara. T.O. acknowledges NSF funding (DEB-2224662, DEB-2340606, and DEB-1748133). Additional funding for specific experimental sites within this synthesis paper came from the USDA Forest Service Rocky Mountain Research Station and the USDA Agricultural Research Service; the findings and conclusions are those of the authors and should not be construed to represent any official USDA determination of policy. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the US Government. Funding for soil analyses was provided by the Colorado Agricultural Experiment Station. Additional funding includes grants from the NSF Long Term Ecological Research (LTER) program (DEB-2423861 and DEB-2426143); funding from the NSF LTREB program to O.S. (DEB-2326482); USGS Ecosystems Mission Area; University of New Hampshire's Agricultural Experiment Station; Natural Sciences and Engineering Research Council of Canada; Canadian Foundation for Innovation; and Federal Ministry of Education and Research (BMBF) grants 031B0516C and 031B1067C. The Ciempozuelos DroughtNet site was established with funds from the European Research Council (BIODESERT, ERC grant agreement no. 647038). F.T.M. acknowledges support from King Abdullah University of Science and Technology; A.C. acknowledges support from the US Department of Agriculture National Institute of Food and Agriculture (grant no. 2024-67019-42396); R.P.P. acknowledges support from the Department of Energy (DoE) Environmental System Science (ESS) program (award no. DE-SC0021980); and L.Y., L.B., and V.F.B. acknowledge support FONCyT (PICT 2019-02324), CONICET (PIP 11220210100681CO), and UBA. Additional funding was provided by National Key R&D Program of China (2022YFE0128000 and 2022YFF1300603) and the National Natural Science Foundation of China (32171592 and 32061123005); New Hampshire Agricultural Experiment Station (accessions 1003450 and 1013351), Iola Hubbard Climate Endowment; Department of Biology at Kansas State University and NSF LTER program to Kansas State University (DEB-144048); FWF (Austrian Science Fund, grant P22214-B17), ÖAW (Austrian Academy of Sciences) grant ClimLUC; Swedish Research Council grants 2013-06395 and 2019-04779, Formas grants 2015-10002 and 2023-00361; Federal Ministry of Education and Research (BMBF) grant 01LL1802C; grants from the US NSF LTER (DEB-1234162 and DEB-1831944); DOE Office of Science Early Career Award DE-SC0015898 and NSF Geography & Spatial Studies grant no. BCS-1437591 to D.F.C.; and German Research Foundation (DFG) grant no. DU1-1/1688. K.M.B. acknowledges support from US Bureau of Land Management (grant no. L16AS00178) and California State University Agricultural Research Institute (grant no. 18-06-004), N.E. acknowledges support of iDiv funded by the German Research Foundation (DFG-FZT 118, 202548816); A.C.G. and G.M.W. acknowledges support from Australian National Landcare Program: DigiFarm-A digitally enabled durable agroecosystem, Hermon Slade Foundation; A.H. acknowledges the John Fell Fund, Ecological Continuity Trust, and Patsy Wood Trust; G.L. acknowledges Programa Atração de Jovens Talentos, CAPES/Brasil; R.O. acknowledges Manitoba Beef and Forage Initiatives Inc. and Canadian Agricultural Partnership; D.C., G.P., and J.M.Z. acknowledge support from PI-UNRN 40-C-1226and PIP-CONICET 402; C.P.C. was supported by the French government IDEX-ISITE initiative 16-IDEX-0001 (CAP 20-25); J.M.P.G. was supported by a postdoctoral fellowship of IAI (CRN3005) and ANPCyT (PICT 2827); M.P.S. acknowledges support from SCRDC-AAFC; K.T. and L.v.d.B. were funded by the German Research Foundation (DFG) Priority Program SPP-1803 "EarthShape: Earth Surface Shaping by Biota" (TI 338/14-1) and ANID PIA/ACT 210038; and C.W. was supported by the China Huaneng Group Headquarters Technology Project (HNKJ22-H48). This work has benefited from the equipment and framework of the COMP-HUB and COMP-R Initiatives, funded by the "Departments of Excellence" program of the Italian Ministry for University and Research (MIUR, 2018-2022 and MUR, 2023-2027). Additional funding includes: the Gottfried Wilhelm Leibniz Prize (Ei 862/29-1); DFG Research Grant DU1-1/1688 to M.D.; Australian Research Council grants (DP150104199 and DP190101968); Norwegian Research Council project (255090); Strategic Environmental Research and Development Program RC18-1322; Natural Sciences and Engineering Research Council of Canada; Federal Ministry of Education and Research (BMBF) grants 01LC1821A, 01LC2321A, 01LL1802C, and 01LL1304D; National Key R&D Program of China (2021YFD2200405) and Sino-Europe Cooperation Project (2023YFE0105100); National Key R&D Program of China (2021YFD2200403) and Sino-Europe Cooperation Project (2023YFE0105100); Teshio Experimental Forest Forest Research Station, Hokkaido University, Japan; Federal Ministry of Education and Research (BMBF) grant 01LL1304D; Estonian Research Council (PRG1065) and the Centre of Excellence AgroCropFuture (TK200); the CredoJ and CredoM Drought-Net sites are part of the Great Western Woodlands TERN SuperSite, supported by the Australian Government and the Western Australian Department of Biodiversity, Conservation, and Attractions; Texas A&M Savanna Long-Term Research and Education Initiative (SLTREI), Sid-Kyle Foundation, and Sonora Research Station personnel; Key Projects of Jilin Province Science and Technology Development Plan (20230303006SF); Strategic Priority Research Program of the Chinese Academy of Sciences (grant no. XDA28110303); National Natural Science Foundation of China (32371669); the Science and Technology Talent Project Distinguished Young Scholars of Jilin Province (20240602009RC); National Natural Science Foundation of China (32071627); National Natural Science Foundation of China (31870456) and National Research Foundation grant no. 116262; Generalitat Valenciana, Project R2D–Responses to Desertification (CIPROM/2021/001); Stengl-Wyer Endowment at University of Texas, Austin; and Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, grants 149862 and 185110.We thank CONAF La Campana and the agricultural community Quebrada de Talca for the access to their sites. A.S.M. was supported by the Environment Research and Technology Development Fund (JPMEERF15S11420) of the Environmental Restoration and Conservation Agency of Japan, with additional field support from the Teshio Experimental Forest, Hokkaido University. Further support came from the Advanced Studies of Climate Change Projection Grant, Ministry of Education, Culture, Sports, Science and Technology, Japan (JPMXD0722678534). **Author contributions:** Conceptualization: T.J.O., M.D.S., S.L.C., A.K.K., J.S.D., O.S., S.M.M., Q.Y., M.L.A., A.J., R.P.P., A.J.F., H.Ab.; Data curation: K.D.W., T.J.O., A.F.J.; Formal analysis and visualization: T.J.O., M.I.A., M.L.A., M.T.H., M.C.H., I.J.S., P.W., K.D.W., A.C., A.J.F., L.A.G., N.J.L., S.P.L.; Writing – original draft: T.J.O., M.D.S., S.L.C., A.K.K., J.S.D., S.M.M., Q.Y.; Writing – review & editing: O.S., M.L.A., A.C., M.T.H., M.C.H., I.J.S., M.E.L., F.T.M., S.A.P., A.J.F., L.A.G., J.A., H.Ab., S.B., M.B., T.L.B., K.H.B., K.B., L.B., E.T.B., K.M.B., J.C., J.F.C., M.C., M.H.C., A.C.C., M.V.C., A.L.C., D.F.C., P.D., L.H.D., M.D., N.E., T.G.W.F., A.C.G., Y.H., H.A.L.H., D.H., F.I., S.E.J., S.E.K., A.Kub., A.Kul., E.G.L., S.P.L., A.L., F.C.L., A.S.M., U.N.N., R.O.-H., R.C.O., G.R.O., B.B.O., R.O., J.P., P.L.P., C.P.-C., J.M.P.-G., S.M.P., S.C.R., C.R., D.W.R., E.W.S., A.S., M.T., T.T., K.T., A.V., L.v.d.B., S.W., G.M.W., C.W., J.L.W., M.Z.; Validation: T.J.O., M.I.A., M.C.H., I.J.S., F.W., K.D.W.; Investigation: T.J.O., M.D.S., S.L.C., J.S.D., O.S., S.M.M., M.C.H., K.D.W., A.J., M.E.L., F.T.M., S.A.P., L.Y., Q.Y., L.A.G., H.Ab., M.A., J.A., A.I.A., H.As., H.Au., S.B., M.B., D.C.B., A.B., T.L.B., K.H.B., K.B., I.Be., L.B., I.Bl., V.F.B., E.T.B., E.W.B., C.M.B., K.M.B., J.F.C., D.A.C., M.C., C.N.C., K.C., M.C.-G., M.H.C., S.X.C., J.C., A.C.C., A.C.C., A.L.C., D.F.C., S.D., P.D., L.H.D., M.D., N.E., T.G.W.F., F.A.F., D.G., S.V.H., Y.H., A.H., H.A.L.H., D.H., F.I., S.E.J., Y.K., E.F.K., S.E.K., J.K., A.I.K., A.Kub., A.Kul., E.G.L., K.S.L., S.L., A.Li., S.L., G.L., A.Lo., J.L., F.C.L., A.V.M., C.D.M., M.V.M., A.S.M., E.M., G.S.N., U.N.N., R.O.-H., R.C.O., R.Og., G.R.O., R.Ot., M.P., J.Pa., J.Pe., P.L.P., D.S.P., G.P., A.P., C.P.-C., J.M.P.-G., L.W.P., R.M.P., C.P.-R., S.M.P., Y.P., G.R., D.A.R., W.E.R., C.R., A.M.S., B.A.S., M.P.S., E.W.S., R.S., B.S., L.S., A.S., R.J.S., M.St., M.Su., M.T., T.J.T., P.T., K.T., M.T., M.A.V., L.v.d.B., V.V., L.G.V., S.W., C. Wei, C.Wer., G.W., J.L.W., A.A.W., C.X., X.Y., A.L.Y., P.Y., J.M.Z., M.Z., J.Z., X.Z.; Supervision and resources: T.J.O., M.D.S., S.L.C., A.K.K., O.S., S.M.M., L.H.F., A.J., M.E.L., F.T.M., R.P.P., S.A.P., L.Y., Q.Y., M.A., J.A., H.As., H.Au., S.B., M.B., T.L.B., K.H.B., V.F.B., E.T.B., C.M.B., K.M.B., J.F.C., C.N.C., J.C., M.V.C., D.F.C., S.A.O.C., P.D., M.D., N.E., A.C.G., Y.H., A.H., H.A.L.H., F.I., S.E.J., S.E.K., J.K., A.I.K., A.K., E.G.L., K.S.L., S.L., A.L., J.L., A.V.M., C.D.M., .D.B.M., A.S.M., G.S.N., U.N.N., R.C.O., G.R.O., B.B.O., R.O., M.P., J.Pa., J.Pe., P.L.P., G.P., A.P., C.P.-C., V.D.P., J.M.P.-G., R.M.P., S.M.P., Y.P., S.C.R., W.E.R., C.R., D.W.R., M.P.S., M.S.-L., E.W.S., L.S., A.S., R.J.S., M.S., W.S., M.T., P.T., K.T., M.A.V., A.V., V.V., G.M.W., C.W., J.L.W., A.A.W., J.L.Y., A.L.Y., M.Z., J.Z. ; Project administration: M.D.S., O.S., R.P.P., S.L.C., A.K.K., J.S.D., C.B., L.H.F., A.J., M.E.L., Y.L., F.T.M., S.A.P., L.Y., Q.Y., H.A. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All code and derived data are publicly available on Dryad (*49*).
License information: Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ads8144
Materials and Methods; Figs. S1 to S6; Tables S1 to S10; References (*51*–*77*);
MDAR Reproducibility Checklist

Submitted 31 August 2024; resubmitted 7 December 2024; accepted 14 August 2025
10.1126/science.ads8144

ORGANIC CHEMISTRY

Synthesis of triple stranded porphyrin nanobelts

Arnau Rodríguez-Rubio[1]†, He Zhu[1]†, Ka Man Cheung[1]†, Igor Rončević[2], Lene A. Gödde[1], Janko Hergenhahn[1], Joshua L. Field[1], Prakhar Gupta[1], Wojciech Stawski[3], Henrik Gotfredsen[1], Joseph Straw[4], Matthew Edmondson[4], James N. O'Shea[4], Alex Saywell[4]*, Harry L. Anderson[1]*

Molecular nanobelts are fascinating analogs of carbon nanotubes. Their rigid geometries and strongly coupled π-electrons have the potential to generate a wave function resembling that of a quantum ring. Here, we report the synthesis of triple stranded nanobelts consisting of 8 to 12 edge-fused porphyrin units with diameters of 21 to 32 angstroms. We synthesized these nanobelts by nickel-mediated coupling of *meso*-bromoporphyrins to form singly linked nanorings, followed by oxidation with gold(III) chloride. Experimental ^{1}H nuclear magnetic resonance spectra, supported by computational simulations, revealed that belts containing odd numbers of porphyrins, with circuits of 90 or 110 π-electrons, display global aromatic ring currents, whereas even-numbered belts, with 80, 100, or 120 π-electrons, are globally antiaromatic.

Single-walled carbon nanotubes adopt three main structural types: zigzag, armchair, and chiral. Very short hydrogen-terminated sections of these tubes are known as carbon nanobelts (*1*–*3*). Hypothetical benzenoid nanobelts have been discussed for decades (*4*), but until recently, all attempts at synthesizing them have failed, owing to their high strain and reactivity. Now, several viable synthetic approaches have been developed (*5*–*11*), leading to intense interest in their potential as optoelectronic materials (*12*, *13*) and as hosts for supramolecular recognition (*8*, *11*). The simplest type of benzenoid nanobelts, and the first to be considered theoretically, are the [*n*]cyclacenes (Fig. 1A) (*1*, *4*). These simple zigzag belts are calculated to have open-shell singlet biradical ground states (*14*), reflecting their lack of Clar sextets (*15*), and this property seems to make them too reactive to isolate. So far, they have only been detected by mass spectrometry (*16*). The biradical character of [*n*]cyclacenes can be suppressed by building in extra benzene rings along the rims of the belt, so that every carbon atom of the π-system becomes part of a Clar sextet (Fig. 1B) (*9*, *10*). One of the most interesting features of a molecular nanobelt is extreme electronic delocalization, potentially making the electronic wave function resemble that of a particle on a ring (i.e., a quantum ring) (*17*, *18*), yet paradoxically, all reported syntheses of benzenoid nanobelts (e.g., Fig. 1, C and D) target nanobelts with many Clar sextets, thus reducing electronic delocalization and increasing chemical stability. All benzenoid nanobelts constructed from six-membered rings are alternant hydrocarbons. Including five-membered rings results in a nonalternant structure and tends to increase the electronic delocalization (*19*–*21*).

Here, we present the synthesis of a family of edge-fused porphyrin nanobelts (***cf*-PN**, where *N* is the number of porphyrin units in the nanobelt; Fig. 1E). The presence of heteroatoms and five-membered rings in these nanobelts results in highly delocalized nonalternant structures. Furthermore, these porphyrin nanobelts are triple stranded (meaning that at least three bonds must be cut to cleave the ring), which amplifies the electronic coupling around the circumference. Previously reported nonporphyrin nanobelts have been double stranded (*1*–*3*, *22*). The electronic delocalization in the neutral edge-fused porphyrin nanobelts is evidenced in their ^{1}H nuclear magnetic resonance (NMR) spectra, which reveal strong global ring currents. Nanobelts with odd and even numbers of porphyrin units have diatropic (aromatic) and paratropic (antiaromatic) global ring currents, respectively, in keeping with Hückel's $4n + 2$ rule and with previously published theoretical predictions (*23*). Another manifestation of the strong global π-conjugation is that the electronic absorption spectra of the neutral nanobelts extend far into the infrared, with intense peaks in the region 1500 to 2000 nm. Previously reported porphyrin-based nanobelts have bridging units separating the porphyrins, rather than direct porphyrin-porphyrin coupling; do not exhibit global ring currents (*22*, *24*–*26*); and do not have such red-shifted absorption spectra.

Nanoring synthesis by nickel-mediated coupling

The edge-fused porphyrin nanobelts reported here are cyclic analogs of the porphyrin nanoribbons pioneered by Tsuda and Osuka (*27*). Previous work has shown that these linear ribbons exhibit exceptionally wirelike charge transport (*28*–*30*), implying that they are promising components for the construction of molecular quantum rings (*18*). Linear porphyrin nanoribbons are synthesized by oxidative fusion of singly linked porphyrin chains (*27*–*32*), so the obvious precursors to edge-fused porphyrin nanobelts are singly linked porphyrin nanorings (***cs*-PN**). We explored various template-directed strategies for preparing the required singly linked rings without success (*33*–*35*). Then, much to our surprise, we found that they are formed serendipitously, together with linear polymers, during the Yamamoto coupling of nickel(II) 5,15-dibromoporphyrins (*28*) (Fig. 2).

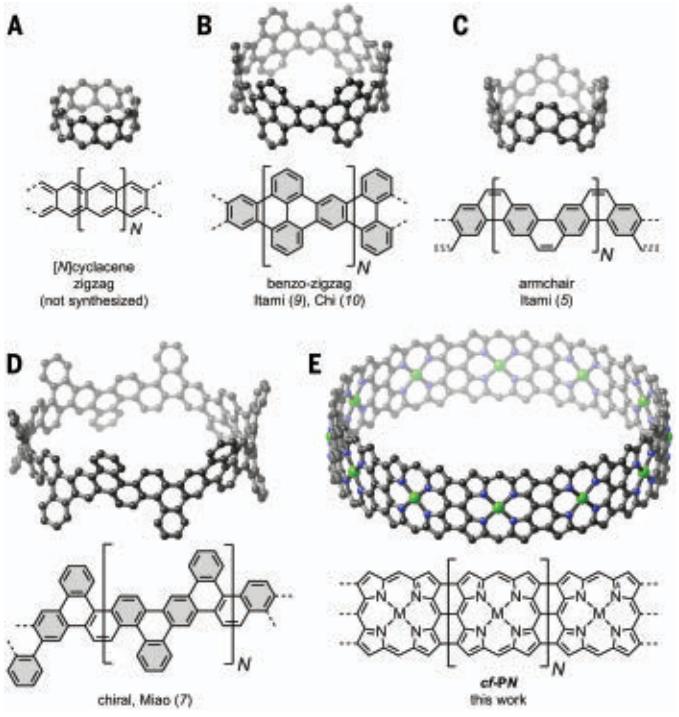


Fig. 1. Comparison of molecular nanobelts. (**A**) [*N*]cyclacene, (**B**) benzo[*N*]cyclacene, (**C**) armchair nanobelt, (**D**) chiral nanobelt, and (**E**) an edge-fused porphyrin nanobelt. The projections of the three-dimensional structures are all shown on the same scale. Clar sextets are shaded gray in (B) to (D).

[1]Department of Chemistry, Oxford University, Chemistry Research Laboratory, Oxford, UK. [2]Department of Chemistry, The University of Manchester, Manchester, UK. [3]Institute of Organic Chemistry II and New Materials, Ulm University, Ulm, Germany. [4]School of Physics and Astronomy, University of Nottingham, Nottingham, UK. ***Corresponding author. Email: harry.anderson@chem.ox.ac.uk (H.L.A.); alex.saywell@nottingham.ac.uk (A.S.)** †These authors contributed equally to this work.

We carried out the Yamamoto coupling reaction shown in Fig. 2, starting from a dibromo nickel(II) porphyrin monomer **P1Br$_2$** or a linear singly linked oligomer ***ls*-PMBr$_2$** (where *M* is the number of porphyrin units) to obtain the cyclic singly linked oligomers ***cs*-PN** with a distribution of ring sizes. These reactions were performed using porphyrins with three different aryl solubilizing groups (OOct, *t*Bu or MesBu, as defined in Fig. 2). The singly linked rings ***cs*-PN** were separated by gel permeation chromatography (GPC) and isolated for N = 6 to 15. (Small amounts of larger rings up to N = 24 were also detected but have not yet been fully characterized.) Cyclic oligomers have longer GPC retention times than linear chains with the same number of porphyrin units (*36*), and the first evidence for the cyclic nature of ***cs*-PN** came from a combination of GPC and mass spectrometry. For example, the GPC trace of a crude reaction mixture from the cyclooligomerization of ***ls*-P2$_{OOct}$Br$_2$** is shown in Fig. 3A, together with the traces of the isolated products ***cs*-P6$_{OOct}$**, ***cs*-P8$_{OOct}$**, ***cs*-P10$_{OOct}$**, and ***cs*-P12$_{OOct}$**. The isolated yields of the cyclic oligomers are plotted as a function of ring size, for various precursors, in Fig. 3B. The small number of signals in the ^{1}H NMR spectra of ***cs*-PN** oligomers was consistent with the formation of highly symmetric cyclic molecules, without end groups, and this was confirmed by scanning tunneling microscopy (STM). For example, Fig. 3C shows an STM image of ***cs*-P8$_{OOct}$** recorded under vacuum after electrospray deposition onto a gold surface [octyloxy-substituted porphyrin polymers are compatible with electrospray deposition and STM imaging on gold surfaces (*34*–*36*)]. Steric repulsion between β-hydrogen atoms results in a severe twist between neighboring porphyrin units in singly *meso-meso*–linked oligomers (*37*), which explains why the STM image of ***cs*-P8$_{OOct}$** has fourfold (rather than eightfold) symmetry [see supplementary materials (SM) and figs. S24 to S26 for more images of ***cs*-P8$_{OOct}$**, ***cs*-P12$_{OOct}$**, and ***cs*-P18$_{OOct}$**].

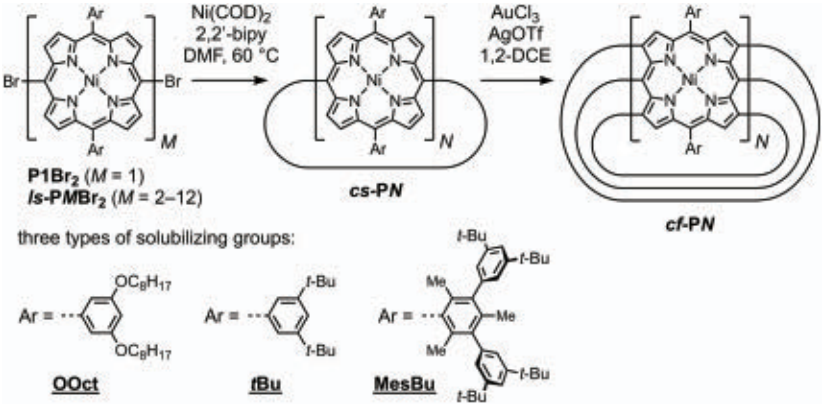


Fig. 2. Synthetic route to porphyrin nanobelts. *M* and *N* are the numbers of porphyrin units in the starting material and rings, respectively. Solubilizing groups are denoted with subscripts; for example, the three monomers have compound codes **P1$_{OOct}$Br$_2$**, **P1$_{tBu}$Br$_2$**, and **P1$_{MesBu}$Br$_2$**. The prefixes ***ls***, ***cs***, and ***cf*** denote linear singly linked, cyclic singly linked, and cyclic fused, respectively. Ar, aryl; 2,2′-bipy, 2,2′-bipyridine; COD, 1,5-cyclooctadiene; 1,2-DCE, 1,2-dichloroethane; DMF, dimethylformamide; OTf, trifluoromethanesulfonate.

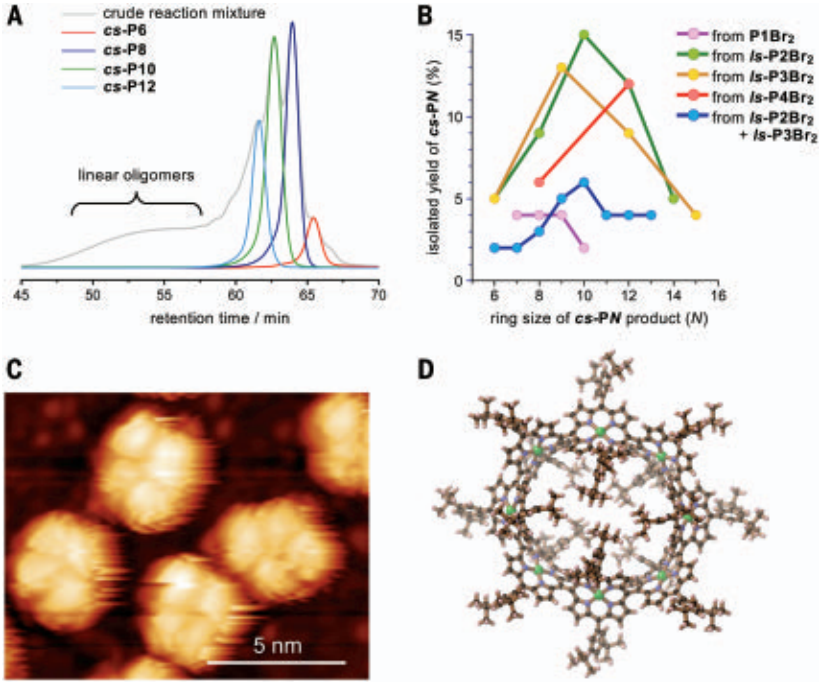


Fig. 3. Synthesis and characterization of singly linked porphyrin nanorings. (**A**) GPC trace of the crude reaction mixture from Yamamoto coupling of the dibromodimer ***ls*-P2$_{OOct}$** superimposed on traces of isolated cyclic oligomers for N = 6, 8, 10, and 12. (**B**) Isolated yields of ***cs*-PN$_{OOct}$** (N = 6 to 15) obtained from a range of different starting materials. (**C**) STM image of ***cs*-P8$_{OOct}$** on a Au(111) surface, deposited by electrospray (imaging parameters: sample bias = −2 V, set-point current = 9 pA). (**D**) Single-crystal x-ray structure of ***cs*-P8$_{tBu}$**, showing one of two distinct molecular units in the crystal.

We grew suitable crystals of ***cs*-P8$_{tBu}$** for single-crystal x-ray crystallography by diffusion of acetone vapor into a solution of the macrocycle in 1,2-dichlorobenzene (see SM for details). There are two crystallographically distinct ***cs*-P8$_{tBu}$** units in the crystal, both of which have similar conformations and one of which is shown in Fig. 3D. The mean Ni–Ni diameter is 20.65(0.66) Å. As expected, the macrocycle has a highly twisted geometry, and neighboring porphyrin units are almost orthogonal. The average angle between the 24-atom planes of neighboring porphyrins is 84.4(2.7)°, and the angle between the porphyrin units and the mean plane of the eight Ni centers is 48.6(6.2)°.

It is surprising that Yamamoto coupling of 5,15-substituted nickel(II) porphyrins gives cyclic oligomers in the absence of any template. Coupling the corresponding zinc(II) porphyrin derivatives under identical conditions only yields linear oligomers (see SM and fig. S9). One reason why nickel(II) porphyrins favor the formation of cyclic products is that they tend to adopt nonplanar geometries because the N-Ni bond length is too short for the cavity of the porphyrin (*38*). This distortion of the porphyrin units reduces the strain in ***cs*-PN** rings when metalated with Ni rather than Zn, as confirmed by density functional theory (DFT) calculations (see SM and fig. S30). For example, we calculated that ***cs*-P16** is almost unstrained when metalated with Ni (strain energy 4.9 kJ mol^{-1}), whereas it has a strain energy of 116.8 kJ mol^{-1} when metalated with Zn.

Another factor that probably favors formation of cyclic products is that nickel(II) porphyrins are electron deficient (*39*), which is expected to retard the reductive elimination step in Yamamoto coupling (*40*). This could lead to the formation of cyclic C_{meso}-Ni-C_{meso}–linked intermediates, in which all the C-Br bonds are replaced by C-Ni-C links before reductive elimination occurs (fig. S31). In other words, metalation of the porphyrins with nickel favors a mechanism in which macrocyclization occurs before reductive elimination. Similar organonickel intermediates have been used in the synthesis of cycloparaphenylenes consisting of electron-deficient fluorinated *para*-phenylene units (*41*) and in the synthesis of

a nanotube end cap (*42*). Strong support for this hypothesis comes from the observation that coupling ***ls*-PMBr$_2$** only gives ***cs*-PN** with $N > M$. For example, ***cs*-P6$_{OOct}$** can be prepared by coupling ***ls*-P2$_{OOct}$Br$_2$** but not from ***ls*-P6$_{OOct}$Br$_2$**, demonstrating that the linear hexamer ***ls*-P6$_{OOct}$Br$_2$** is not an intermediate in the formation of the cyclic hexamer ***cs*-P6$_{OOct}$** from ***ls*-P2$_{OOct}$Br$_2$** (see fig. S7B). Similarly, coupling of ***ls*-P12$_{OOct}$Br$_2$** does not give ***cs*-P12$_{OOct}$** (fig. S8). Reversible formation of C_{meso}-Ni-C_{meso} links might allow equilibration between linear and cyclic intermediates, which could also favor macrocyclization.

Oxidation to nanobelts and global aromaticity

Oxidation of the singly linked rings ***cs*-PN** to the triply linked nanobelts ***cf*-PN** was achieved with gold(III) chloride and silver triflate in around 95% yield using reaction conditions developed for the synthesis of linear porphyrin ribbons (*28*, *30*, *43*). The role of the silver salt is to suppress formation of chlorinated by-products (*43*). This fusion reaction is accompanied by a pronounced change in the ultraviolet–visible–near-infrared (UV-vis-NIR) absorption spectra, and the products exhibit intense absorption bands at 1500 to 2000 nm (Fig. 4). We tested this chemistry with all three aryl (Ar) groups shown in Fig. 2, and similar UV-vis-NIR spectra were recorded in all three cases; however, sharp well-resolved ^{1}H NMR spectra were only obtained for ***cf*-PN$_{MesBu}$**, so we focused on this family of nanobelts ($N = 8$ to 12). We attribute the broadness of the ^{1}H NMR spectra of ***cf*-PN$_{OOct}$** and ***cf*-PN$_{tBu}$** to dynamic rotation of the aryl groups (which is blocked by the *ortho*-methyl groups in ***cf*-PN$_{MesBu}$**) and aggregation (which is hindered by the bulk of the MesBu groups). The ^{1}H NMR spectra of the five ***cf*-PN$_{MesBu}$** nanobelts (Fig. 5) were assigned using a combination of two-dimensional NMR techniques (see SM for details). It was not possible to distinguish between mesityl methyl groups facing inside (Me_{in}) or outside (Me_{out}) of the nanobelts using nuclear Overhauser effect correlations. We assigned these resonances by assuming that the

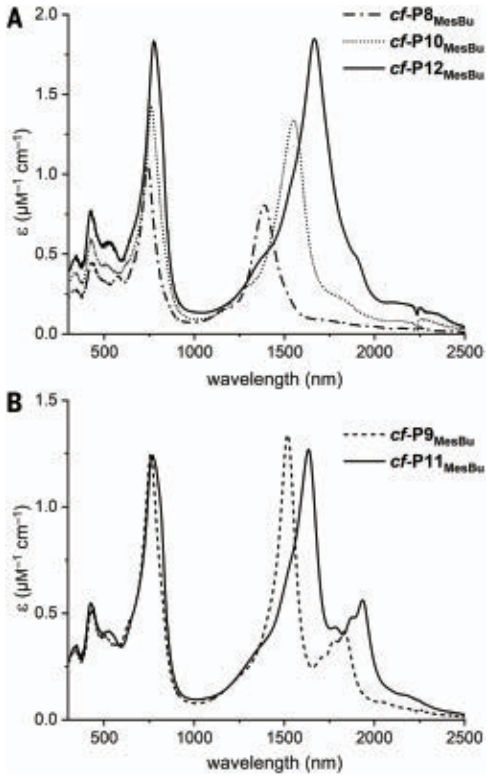


Fig. 4. UV-vis-NIR spectra of nanobelts recorded in toluene-d_8. (**A**) Antiaromatic belts with even numbers of porphyrin units. (**B**) Aromatic belts with odd numbers of porphyrin units.

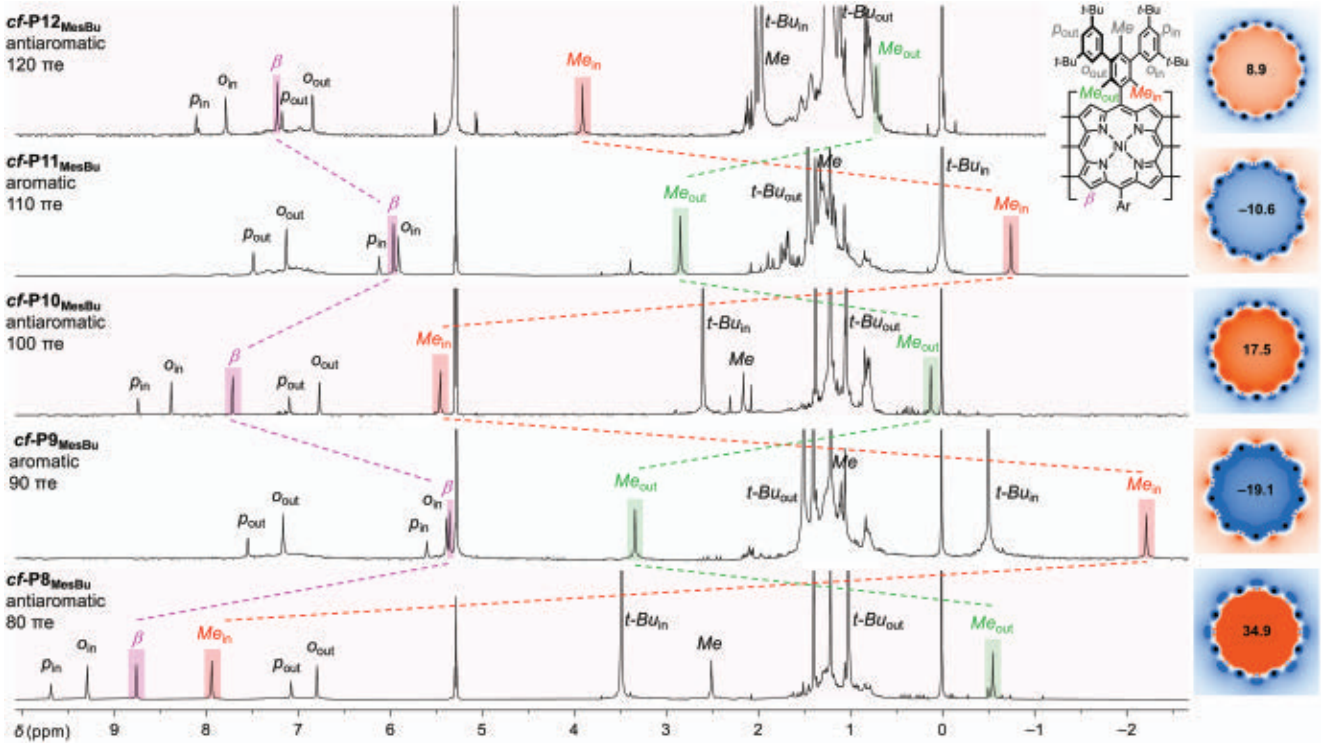


Fig. 5. ^{1}H NMR spectra of porphyrin nanobelts *cf*-P8$_{MesBu}$ to *cf*-P12$_{MesBu}$ recorded in CD_2Cl_2/CS_2 (at 400 MHz, 298 K). Plots of the NICS in the plane of the porphyrin for each ring size are shown on the right [$NICS(0)_{zz}$ from DFT calculations at the OX-B3LYP/def2-SVP level of theory (*23*). Numbers denote the NICS value at the center of the ring.] ppm, parts per million.

signal most shielded, or most deshielded, by the global ring current is Me_{in} (*44*). This assignment was fully confirmed by the excellent correlation between experimental chemical shifts and those calculated using DFT (fig. S15). The ^{1}H NMR spectra show marked changes in the chemical shifts of the Me_{in}, Me_{out}, and β signals, which alternate for odd and even numbers of porphyrin units, indicating that ***cf*-P9$_{MesBu}$** and ***cf*-P11$_{MesBu}$** are globally aromatic, whereas ***cf*-P8$_{MesBu}$**, ***cf*-P10$_{MesBu}$**, and ***cf*-P12$_{MesBu}$** are globally antiaromatic, as predicted (*23*), consistent with the nucleus-independent chemical shift (NICS) plots shown on the right side of Fig. 5. The ring current susceptibilities of the nanobelts with $N = 8$ to 12 were estimated from the experimental ^{1}H NMR shifts in Me_{in} and Me_{out} using the Biot-Savart law (*18*), giving values of 41.9, –29.8, 31.3, –20.0, and 17.2 nA T^{-1}, respectively [compared with –11.8 nA T^{-1} in benzene (*45*)]. The integrated bond currents in ***cf*-P8** and ***cf*-P9**, calculated using SYSMOIC (*46*) (fig. S16), are similar in magnitude to ring currents from Biot-Savart law analysis, with about 90% of the current flowing through the C_{meso}-C_{meso} bonds, as predicted for the corresponding zinc complexes (*23*).

The UV-vis-NIR absorption spectra of aromatic compounds generally feature sharp intense bands, whereas antiaromatic compounds tend to give broader absorption spectra that tail into the NIR region (*47*). We did not observe a notable difference between the absorption spectra of the nanobelts in the series ***cf*-P8$_{MesBu}$** to ***cf*-P12$_{MesBu}$**, but there is an alternation in the shape of the NIR spectra between even and odd nanobelts: The aromatic belts ***cf*-P9$_{MesBu}$** and ***cf*-P11$_{MesBu}$** exhibit a pattern of sharp shoulder bands at 1700 to 1900 nm, which is absent in the antiaromatic belts ***cf*-P8$_{MesBu}$**, ***cf*-P10$_{MesBu}$**, and ***cf*-P12$_{MesBu}$** (Fig. 4).

Conclusions

We presents the synthesis of molecular nanobelts consisting of 8 to 12 edge-fused porphyrin units. The ^{1}H NMR spectra of these neutral compounds reveal that they sustain global ring currents that are diatropic (aromatic) for odd numbers of porphyrins and paratropic (antiaromatic) for even numbers of porphyrins. These findings are in stark contrast with the behavior of previously reported nanobelts (and butadiyne-linked porphyrin nanorings), which only exhibit global ring currents when oxidized or reduced (*18*, *20*, *21*, *44*). The observation of global (anti)aromatic ring currents in neutral porphyrin nanobelts demonstrates that the electronic wave function resembles, to some extent, that of a particle on a ring (i.e., a quantum ring) (*17*, *18*), which implies that charge transport through these molecules can be controlled by quantum interference. For example, theoretical work indicates that ***cf*-P*N*** nanobelts could function as Aharonov-Bohm–type interferometers (*48*), enabling transport to be controlled by applying a magnetic field.

REFERENCES AND NOTES

1. Q.-H. Guo, Y. Qiu, M.-X. Wang, J. F. Stoddart, *Nat. Chem.* **13**, 402–419 (2021).
2. D. Imoto, A. Yagi, K. Itami, *Precis. Chem.* **1**, 516–523 (2023).
3. R. Zhang, D. An, J. Zhu, X. Lu, Y. Liu, *Adv. Funct. Mater.* **33**, 2305249 (2023).
4. E. Heilbronner, *Helv. Chim. Acta* **37**, 921–935 (1954).
5. G. Povie, Y. Segawa, T. Nishihara, Y. Miyauchi, K. Itami, *Science* **356**, 172–175 (2017).
6. G. Povie, Y. Segawa, T. Nishihara, Y. Miyauchi, K. Itami, *J. Am. Chem. Soc.* **140**, 10054–10059 (2018).
7. K. Y. Cheung *et al.*, *Chem* **5**, 838–847 (2019).
8. H. M. Bergman, G. R. Kiel, R. C. Handford, Y. Liu, T. D. Tilley, *J. Am. Chem. Soc.* **143**, 8619–8624 (2021).
9. K. Y. Cheung, K. Watanabe, Y. Segawa, K. Itami, *Nat. Chem.* **13**, 255–259 (2021).
10. Y. Han, S. Dong, J. Shao, W. Fan, C. Chi, *Angew. Chem. Int. Ed.* **60**, 2658–2662 (2021).
11. Z. Xia, S. H. Pun, H. Chen, Q. Miao, *Angew. Chem. Int. Ed.* **60**, 10311–10318 (2021).
12. J. Lin *et al.*, *Sci. Adv.* **8**, eade4692 (2022).
13. M. Ball *et al.*, *Acc. Chem. Res.* **52**, 1068–1078 (2019).
14. Z. Chen *et al.*, *Org. Lett.* **9**, 5449–5452 (2007).
15. E. Clar, *The Aromatic Sextet* (Wiley, 1972).
16. T.-H. Shi, Q.-H. Guo, S. Tong, M.-X. Wang, *J. Am. Chem. Soc.* **142**, 4576–4580 (2020).
17. M. Solà, F. M. Bickelhaupt, *J. Chem. Educ.* **99**, 3497–3501 (2022).
18. M. Jirásek, H. L. Anderson, M. D. Peeks, *Acc. Chem. Res.* **54**, 3241–3251 (2021).
19. C. Liu *et al.*, *Chem* **4**, 1586–1595 (2018).
20. X. Lu *et al.*, *J. Am. Chem. Soc.* **141**, 5934–5941 (2019).
21. Y. Han, S. Wu, K. Y. S. Khoo, C. Chi, *Nat. Synth.* **4**, 947–955 (2025).
22. R. Haver, H. L. Anderson, *Helv. Chim. Acta* **102**, e1800211 (2019).
23. M. Vitek, J.-R. Deng, H. L. Anderson, I. Rončević, *ACS Nano* **19**, 1405–1411 (2025).
24. J. Song, N. Aratani, H. Shinokubo, A. Osuka, *J. Am. Chem. Soc.* **132**, 16356–16357 (2010).
25. P. Neuhaus, A. Cnossen, J. Q. Gong, L. M. Herz, H. L. Anderson, *Angew. Chem. Int. Ed.* **54**, 7344–7348 (2015).
26. S. Xue, D. Kuzuhara, N. Aratani, H. Yamada, *Org. Lett.* **21**, 2069–2072 (2019).
27. A. Tsuda, A. Osuka, *Science* **293**, 79–82 (2001).
28. H. Zhu *et al.*, *Angew. Chem. Int. Ed.* **64**, e202417429 (2025).
29. S. M. Kopp *et al.*, *Chem* **10**, 3595–3606 (2024).
30. J.-R. Deng, M. T. González, H. Zhu, H. L. Anderson, E. Leary, *J. Am. Chem. Soc.* **146**, 3651–3659 (2024).
31. T. Ikeda, N. Aratani, A. Osuka, *Chem. Asian J.* **4**, 1248–1256 (2009).
32. T. Tanaka, A. Osuka, *Chemistry* **24**, 17188–17200 (2018).
33. J. M. Van Raden *et al.*, *Angew. Chem. Int. Ed.* **63**, e202400103 (2024).
34. M. A. Majewski *et al.*, *Angew. Chem. Int. Ed.* **62**, e202302114 (2023).
35. H. Gotfredsen *et al.*, *Nat. Chem.* **14**, 1436–1442 (2022).
36. D. V. Kondratuk *et al.*, *Nat. Chem.* **7**, 317–322 (2015).
37. N. Aratani *et al.*, *Chemistry* **11**, 3389–3404 (2005).
38. X.-Z. Song *et al.*, *J. Am. Chem. Soc.* **118**, 12975–12988 (1996).
39. D. G. Davis, in *The Porphyrins*, **vol. 5**, D. Dolphin, Ed. (Academic Press, 1978), pp. 127–152.
40. T. Yamamoto, Y. Murakami, M. Abla, *Chem. Lett.* **28**, 419–420 (1999).
41. H. Shudo *et al.*, *Nat. Commun.* **13**, 3713 (2022).
42. D. Myśliwiec, M. Kondratowicz, T. Lis, P. J. Chmielewski, M. Stępień, *J. Am. Chem. Soc.* **137**, 1643–1649 (2015).
43. A. K. Sahoo *et al.*, *Org. Lett.* **8**, 4141–4144 (2006).
44. M. Jirásek, M. Rickhaus, L. Tejerina, H. L. Anderson, *J. Am. Chem. Soc.* **143**, 2403–2412 (2021).
45. H. Fliegl, D. Sundholm, S. Taubert, J. Jusélius, W. Klopper, *J. Phys. Chem. A* **113**, 8668–8676 (2009).
46. G. Monaco, F. F. Summa, R. Zanasi, *J. Chem. Inf. Model.* **61**, 270–283 (2021).
47. J. Oh, Y. M. Sung, Y. Hong, D. Kim, *Acc. Chem. Res.* **51**, 1349–1358 (2018).
48. C. Y. Cheng *et al.*, *Chem. Sci.* **16**, 4392–4401 (2025).
49. I. Rončević, H. L. Anderson, J. Hergenhahn, Supplementary computational data for triple stranded porphyrin nanobelts. Zenodo (2025); https://doi.org/10.5281/zenodo.15675187.
50. K. M. Cheung, A. Rodríguez-Rubio, H. Zhu, H. L. Anderson, Supplementary experimental data for "Synthesis of triple stranded porphyrin nanobelts." Zenodo (2025); https://doi.org/10.5281/zenodo.16738579.

ACKNOWLEDGMENTS

We thank M. Vitek for discussions. **Funding:** This work was funded by European Research Council (ERC) grant 885606, ARO-MAT (H.L.A., A.R.-R., K.M.C., J.H., P.G., and W.S.); The Royal Society for a University Research Fellowship (A.S.); UK Research and Innovation (UKRI) Horizon Europe Guarantee MSCA Postdoctoral Fellowship ElDelPath, EP/X030075/1 (I.R., H.L.A.); and the Centre for Doctoral Training in Synthesis for Biology and Medicine for a studentship, generously supported by GlaxoSmithKline, MSD, Syngenta, and Vertex (J.L.F.). **Author contributions:** Conceptualization: H.L.A., A.R.-R., H.Z., K.M.C.; First serendipitous observation of Yamamoto cyclization: H.Z.; Synthetic experimental work and spectroscopy: A.R.-R., H.Z., K.M.C., J.L.F., H.G.; Theoretical analysis and computational simulations: I.R., J.H.; Biot-Savart calculations: L.A.G.; X-ray crystallography: P.G., W.S.; Scanning probe microscopy: J.S., M.E., J.O., A.S.; Writing: All authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** X-ray data are available free of charge from the Cambridge Crystallographic Data Centre under accession numbers CCDC-2463702 and CCDC-2463698. All other data are plotted in the main text and the supplementary materials and are deposited in numerical format in Zenodo (*49*,*50*). Detailed methods to prepare the reported compounds are given in the supplementary materials.

License information: Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse.

This research was funded in whole or in part by UKRI (grant EP/X030075/1), a cOAlition S organization, and by the ERC (grant 885606); as required, the author will make the Author Accepted Manuscript (AAM) version available under a CC BY public copyright license.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adz8900
Materials and Methods; Supplementary Text; Figs. S1 to S116; Tables S1 to S14; References (*51*–*61*)

Submitted 18 June 2025; accepted 21 August 2025

10.1126/science.adz8900

ELECTROCHEMISTRY

Durable, pure water–fed, anion-exchange membrane electrolyzers through interphase engineering

Shujin Hou[1,2], Archana Sekar[3], Yang Zhao[1,2], Minkyoung Kwak[1,4], Juhyun Oh[5], Kelvin Kam-Yun Li[1,6,7], Peiyao Wu[1], Ryan T. Hannagan[8,9], Valeria Cartagena[8,9], Anthony C. Ekennia[1,4], Hui Duan[10], Michael J. Zachman[5], Joelle Frechette[1,2], Gregory M. Su[6,7], Balsu Lakshmanan[3], Yushan Yan[10], Thomas F. Jaramillo[8,9], Shannon W. Boettcher[1,2,4*]

Anion-exchange membrane water electrolyzers (AEMWEs) promise scalable, low-cost hydrogen production but are limited by the electrochemical instability of their anode ionomers. We report interphase engineering using inorganic-containing molecular additives that coassemble with ionomer, enabling pure water–fed AEMWEs to operate with a degradation rate <0.5 millivolt per hour at 2.0 amperes per square centimeter and 70°C—a >20-fold durability improvement. Analysis of different additives and ionomers shows that the stabilization mechanism involves cross-links between metal oxo/hydroxo oligomers and ionomers. Under operation, the inorganic additive enriches, forming an interphase near the water-oxidation catalyst that passivates the anode ionomer against continuous degradation while maintaining mechanical integrity and hydroxide conductivity. This additive-based interphase-engineering strategy provides a path to durable AEMWEs that operate without supporting electrolytes and is adaptable across diverse catalysts and ionomers for electrochemical technologies.

Anion-exchange membrane water electrolyzers (AEMWEs) offer electrochemical output-gas compression (*1*, *2*), load following and start-stop operation (*3*, *4*), pure-water feed (*5*), and the possibility of using hydrocarbon ionomer membranes (*1*, *6*), non–precious metal catalysts (*7*, *8*), and low-cost stack components (*1*, *9*). AEMWE is thus a possible disruptive technology to enable H_2 production to outcompete fossil fuel–derived H_2 on cost (*10*, *11*). Although the performance of research-scale and commercial systems has improved, durability remains a major issue, especially when operating without soluble supporting electrolytes as is needed to simplify the balance of plant (*12*) and eliminate shunt currents and reverse polarization (*13*, *14*).

The ionomers used in AEMWEs contain fixed-charge groups, typically quaternary ammonium cations with effective concentrations >1 M. The mobile hydroxide counterions (OH^-) carry the ionic current through the porous electrode and membrane phases (*15*, *16*). Historically, these ionomers have been susceptible to degradation by Hofmann elimination and nucleophilic attack of OH^- on the cationic group or backbone (*10*, *17*, *18*) owing to the more aggressive alkaline environment compared to acidic conditions. Modern ionomer chemical designs mitigate the attack of OH^- using steric and electronic effects (*7*, *19*–*25*). Yet, when these ionomers are used in AEMWEs as the membrane, and within the porous electrodes containing ionomer and catalyst, the durability is poor with degradation rates typically >>1 mV/hour even at a current density of ≤1 A cm^{-2} (*3*, *5*, *13*, *26*–*30*). Such performance is not competitive with proton-exchange membrane water electrolyzers (PEMWEs) that operate at 2 to 4 A cm^{-2} at <2 V and with <10 μV/hour degradation rates (*31*, *32*).

The primary degradation mode in pure water–fed AEMWE is the oxidation of the ionomer within the anode catalyst layer under the locally alkaline conditions in which even the most robust perfluorinated ionomers degrade (*13*, *33*). Ionomer oxidation leads to increased ohmic resistance through the loss of OH^- transport pathways (*34*, *35*), dissolution of transition-metal catalysts driven by decreased local pH as the basic quaternary ammonium groups are degraded (*36*–*38*), and mechanical disintegration and catalyst detachment from compromised binding properties of the ionomer (*36*). Recent prototypes for commercial AEMWE systems circumvent this issue by flooding the anode with a soluble KOH electrolyte that carries the ionic current (*1*, *5*), in addition to the ionomer binder, resulting in a hybrid device that mixes elements of traditional liquid-alkaline and membrane electrolyzers. Even in supporting electrolyte, however, the oxidation rate is nonzero. The supporting electrolyte masks this degradation, particularly over shorter tests, by maintaining conformal OH^- transport to the catalyst (*37*, *39*, *40*). The hybrid approach also negates some advantages of a pure-water feed, including the simplified balance of plant, lower-cost materials, and lack of reverse polarization (*13*). The development of new anode and ionomer interfacial strategies to suppress oxidation of ionomers is thus central to making AEMWEs a disruptive technology.

Li-ion batteries also feature reactive electrodes that degrade organic electrolyte and salts (*41*–*43*). This problem has been solved in commercial Li-ion batteries through a combination of electrolyte additives that decompose on the reactive electrodes to form organic-inorganic solid-electrolyte interphase (SEI) layers that passivate the electrode against continuous electrolyte decomposition (*43*, *44*). The SEI is conductive to Li^+ but blocks electronic current that would otherwise result in catastrophic electrolyte decomposition (*45*). Although the interphase designs will differ, conceptually similar SEI engineering approaches may be much more broadly applicable in stabilizing electrodes in electrochemical technology.

Here we show that metal-alkoxide inorganic additives generate oxo and hydroxo oligomers that can be tuned to coassemble with ionomers during electrode fabrication and form durability-enhancing interphases. The most effective Zr-based additives create inorganic cross-links in the ionomer phase that improve stability of the electrode more than 10-fold. These layers limit interfacial degradation through formation of Zr-oxo/hydroxo–enriched interphases that appear resistant to oxidation and electron transport but have high OH^- conductivity and no detrimental effect on the catalyst's water-oxidation kinetics. With this interphase-engineering strategy, we show a pure water–fed AEMWE operating for >400 hours at 2.0 A cm^{-2} at 70°C degrading at <0.5 mV $hour^{-1}$. The approach is versatile and compatible with many catalysts, metal-alkoxide or other inorganic additives, and ionomers. The work provides a new science-informed pathway to practical AEMWEs that may, with further development, outcompete PEMWEs on performance and traditional liquid-alkaline electrolyzers on cost. This additive and interphase-engineering approach is broadly relevant to enhancing the durability of other ionomer-based electrochemical devices that operate in harsh environments.

Design of additives-ionomer interphase

The chemistry underlying the preparation of durable anodes is shown in Fig. 1, using zirconium propoxide additive, $Zr(OPr)_4$, as an example,

[1]Department of Chemical and Biomolecular Engineering and Department of Chemistry, University of California, Berkeley, Berkeley, CA, USA. [2]Energy Storage and Distributed Resources Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [3]Versogen, Newark, DE, USA. [4]Oregon Center for Electrochemistry, Department of Chemistry and Biochemistry, University of Oregon, Eugene, OR, USA. [5]Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN, USA. [6]Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [7]Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [8]Department of Chemical Engineering, Stanford University, Stanford, CA, USA. [9]SUNCAT Center for Interface Science and Catalysis, SLAC National Accelerator Laboratory, Menlo Park, CA, USA. [10]Department of Chemical and Biomolecular Engineering, Center for Clean Hydrogen, University of Delaware, Newark, DE, USA. ***Corresponding author. Email: boettcher@berkeley.edu**

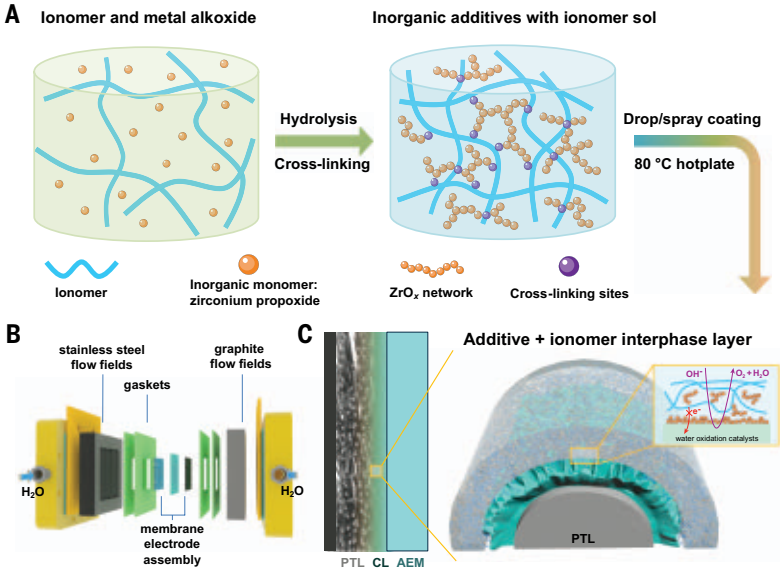


Fig. 1. Engineering a solid-electrolyte interphase in AEMWE devices. (**A**) Preparation of organic-inorganic gel by cross-linking ionomer and zirconium oxo/hydroxy oligomers. (**B**) Schematic of the pure water–fed AEMWE single cell. (**C**) Schematic of the hypothesized additive-ionomer interphase structure located between the ionomer and the anode catalyst. The cylindrical shape represents the stainless-steel fiber that comprises the PTL, which is conformally coated by the nanosheet-shaped catalyst prior to additive-ionomer infilling.

with PiperION anion-exchange ionomer. When the transparent, yellowish ionomer dispersion is added to $Zr(OPr)_4$, the mixture turns milky white (fig. S1) as the metal alkoxides partially hydrolyze, $Zr(OPr)_4 + H_2O \rightarrow Zr(OH)(OPr)_3 + PrOH$ (*46*), and react with the bicarbonate counterions, $Zr(OPr)_4 + HCO_3^- \rightarrow Zr(OPr)_3(CO_3)^- + PrOH$ (*47*). These new species condense to form oligomers, e.g., via $Zr(OPr)_3(OH) + Zr(OPr)_3(OH) \rightarrow Zr(OPr)_3{-}O{-}Zr(OPr)_3 + H_2O$. Simultaneously, the hydrolyzed inorganic species interact with the organic ionomers, forming chemical cross-links (see section on chemical interactions below). The resulting inorganic oligomer–ionomer dispersion is then coated at 80°C onto the anode catalysts (here typically CoO_x-based) that are grown directly on a Ni or steel porous transport layer (PTL). As alcohol and water evaporate, polycondensation is promoted and a cross-linked inorganic-organic hybrid interphase that has improved oxidative resistance and mechanical adhesion is formed (Fig. 1C).

Performance of AEMWEs without soluble supporting electrolyte

Using electrodeposited cobalt hydroxide [e-$Co(OH)_2$] anode catalyst as an example, we optimized the mass ratio of ionomer and $Zr(OPr)_4$ in the dispersion, showing a two-order-of-magnitude reduction in the degradation rate over the first 20 hours compared to ionomer-only coatings (Fig. 2, A and B, and figs. S13A and S15). Beginning-of-test (BOT) voltages decreased with increasing ionomer:$Zr(OPr)_4$ mass ratio, whereas degradation rates followed a U-shaped trend and a minimum of <0.1 mV $hour^{-1}$ over 20 hours at 2.0 A cm^{-2} with a ratio of 23:1. For ionomer:$Zr(OPr)_4$ mass ratios <20, the electrolyzer BOT performance was inferior to that of the ionomer-only system (2.0 A cm^{-2} at 2.17 ± 0.02 V), likely owing to insufficient ionomer in the catalyst layer limiting ionic conductivity and causing an increase in the measured high-frequency resistance (fig. S14). Although lower $Zr(OPr)_4$ content improved BOT cell performance, it compromised durability.

Given the three-dimensional structure and distribution of e-$Co(OH)_2$ on or within the PTL (fig. S16), the method of applying additive-ionomer dispersion has a substantial effect on cell voltage and durability. The lowest degradation rates were achieved by drop casting 1.5 mg cm^{-2} of additives-ionomer on the catalyst-coated PTL, followed by spray coating 1.5 mg cm^{-2} (fig. S13B). Counterintuitively, three-electrode measurements in 0.10 M KOH showed that coating e-$Co(OH)_2$ with the Zr-propoxide additive enhanced the water-oxidation activity by ~8% (fig. S17). The improvement may be due to ZrO_x modifying the catalyst surface chemistry, perhaps increasing hydrophilicity under oxygen evolution reaction conditions (fig. S18) (*48*). Electrochemical surface-area studies in both electrolyzer and three-electrode systems indicate that incorporating the additive into the ionomer layer has a negligible impact on apparent capacitance and effectively maintains the BOT capacitance after 20-hour durability tests (figs. S22 and S23). Online inductively coupled plasma mass spectrometry (ICP-MS) showed similar stability of the e-$Co(OH)_2$ during short durability tests with and without additive (fig. S5), consistent with the hypothesis that the ionomer is the source of instability.

To understand transport through the additives-ionomer interphase layer, we measured the ionic conductivity in the AEMWE (*49*, *50*). We sandwiched a layer of ionomer, or ionomer with additives, between two strips of membrane that provided an ionic interconnect to two reference electrodes exterior to the cell. The voltage across the reference electrodes was measured as a function of current through the electrolyzer. The additives-ionomer film had ~12% higher ionic conductivity than ionomer alone under identical conditions (fig. S19). Similarly, a film containing Zr additive yielded lateral conductivity of ~190 mS cm^{-1} (80°C under 100% relative humidity, fig. S20), higher than that of pristine PAP-TP-85 (~170 mS cm^{-1}) (*19*). We also found that the cell-level voltage losses from porous-electrode electronic and ionic transport (*51*) decreased after a short-term durability test when Zr additive was incorporated in ionomer (fig. S21), consistent with improved OH^- transport. The ZrO_x is hydrophilic, which appears to drive assembly of water channels and thus improve the internal ion-conducting pathways (see section on chemical interactions below). Incorporating optimal inorganic additives thus unexpectedly enhances both catalytic activity for water oxidation and ionic transport to those catalytic sites.

During longer-term measurements at 70°C (Fig. 2B), degradation rates stabilized with an average degradation rate of 0.39 mV $hour^{-1}$ at 1.0 A cm^{-2} over 200 hours, and 0.49 mV $hour^{-1}$ at 2.0 A cm^{-2} over 400 hours, with no abrupt cell failure detected. The average degradation rates were 0.31 and 0.80 mV $hour^{-1}$ at 1.0 and 2.0 A cm^{-2}, respectively, over the first 120 hours. The optimized dispersion with ionomer:$Zr(OPr)_4$ mass ratio of 23 was used with other catalysts, including commercial Co_3O_4 and IrO_x powders, hydrothermally synthesized cobalt hydroxide [h-$Co(OH)_2$]. For the powder samples, the catalyst layer was fabricated by spray-coating inks containing nanopowder catalyst and ionomer, followed by spray-coating 2.0 mg cm^{-2} of additives-ionomer as a top-coat. All catalysts with additives-ionomer exhibited BOT cell voltages that were comparable to those of ionomer-only counterparts and end-of-test (EOT) cell voltages that were significantly lower (Fig. 2C and fig. S15), with degradation rates reduced by at least fivefold using the Zr additive (Fig. 2D).

Microscopic electrocatalyst-interphase-ionomer structure

We studied the structure of the electrode to understand the processes underlying enhanced stability. In the ionomer-only system, electron microscopy and spectroscopy (figs. S28 to S31) showed the continuous loss of ionomer on the catalyst surface, consistent with measured increases of cell voltage, and ohmic and charge-transfer impedances (figs. S32 and S33). Ionomer oxidation and dissolution weaken mechanical adhesion and facilitate particle detachment. For cells with optimal $Zr(OPr)_4$, the additives-ionomer interphase layer remained adhered to the electrocatalyst surface even after 20 hours at 2.0 A cm^{-2} (figs. S34 and S35).

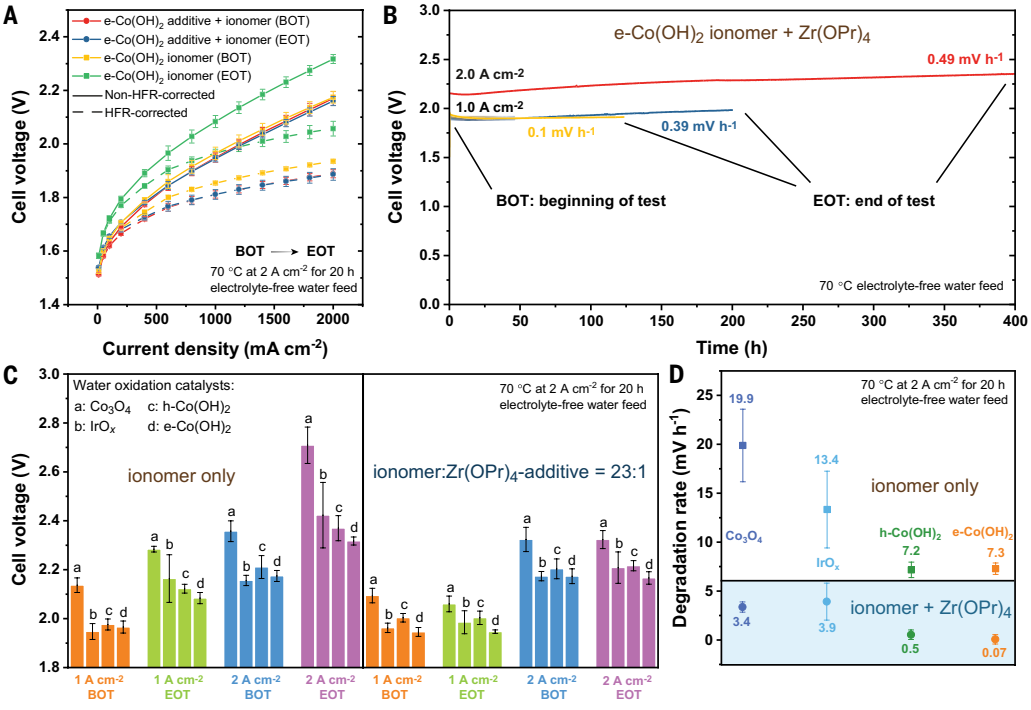


Fig. 2. AEMWE performance when fed pure water. (**A**) Beginning-of-test (BOT) and end-of-test (EOT) polarization curves of AEMWE cells with electrodeposited cobalt hydroxide [e-Co(OH)$_2$] as the anode catalyst layer interfaced with either ionomer or an ionomer-additive mixture for a 20-hour test at 2.0 A cm^{-2} and 70°C. Pure water (18.2 MΩ cm) was used as the feed. The dashed lines represent the polarization curves corrected for high-frequency resistance (HFR), before and after the 20-hour durability tests. (**B**) Chronopotentiometry stability test using e-Co(OH)$_2$ anodes coated with additives-ionomer at 1.0 and 2.0 A cm^{-2}, where the device tests were ended after 123, 200, and 400 hours without device failure. (**C**) BOT and EOT cell voltages at 1.0 and 2.0 A cm^{-2} for different anode catalysts. The left panel shows results for ionomer alone and the right panel with additives. This durability test was carried out at 70°C for 20 hours in pure-water feed. Commercial Co_3O_4, IrO_x, hydrothermally synthesized cobalt hydroxide [h-Co(OH)$_2$] and e-Co(OH)$_2$ were loaded onto stainless-steel porous transport layers as the anode catalysts. (**D**) Degradation rates of AEMWEs using the above four different anode catalysts over 20 hours at a current density of 2.0 A cm^{-2}, each showing marked improvements in durability with the additive-derived interphase.

Low-dose cryo–transmission electron microscopy (cryo-TEM) was used to study the system. Model electrodes were prepared by depositing e-Co(OH)$_2$ onto Au-mesh TEM grids; subsequently coated with ionomer with, and without, Zr(OPr)$_4$ additive; and operated in the AEMWE cell at 2.0 A cm^{-2} (fig. S36). The sample with additive displayed distinct, nanometer-scale, dark features within the ionomer and near the e-Co(OH)$_2$ catalyst surface, which we attributed to amorphous ZrO_x-rich regions (Fig. 3B). The sample without additive showed few high-contrast features (Fig. 3A). Electron energy-loss spectroscopy (EELS) of the e-Co(OH)$_2$-additive-ionomer sample (fig. S40) showed the Zr L-edge, which was not present in the control sample. Spectral features at the N K-edge associated with quaternary ammonium groups changed little for the e-Co(OH)$_2$-additives-ionomer sample after the 20 hours, compared to the e-Co(OH)$_2$-ionomer sample (fig. S6), consistent with the ZrO_x protecting these groups from oxidative damage (fig. S8). Time-of-flight secondary-ion mass spectrometry depth profiles showed a uniform distribution of ZrO_x and ionomer in films coated on a flat iridium-coated silicon control wafer (fig. S41).

Cross-sectional maps (Fig. 3, C to F, and figs. S42 to S45) showed Zr uniformly distributed in the pristine samples. After 1 hour of operation at 2.0 A cm^{-2}, increased Zr was found in the regions next to the electrocatalyst, consistent with oxidative instability of the ionomer near the catalyst leading to its degradation and dissolution. The remaining ZrO_x-rich interphase is hypothesized to be a porous, hydrated, hydroxide-permeable solid that maintains mechanical stability and adhesion while electrically insulating against catastrophic oxidative ionomer damage.

The accumulation of ZrO_x in the interphase appears critical in preventing runaway interfacial oxidative degradation of the ionomer and associated increased resistance and pH decrease at the catalyst. Control experiments in which 0.7 mg cm^{-2} of ionomer was coated on the catalyst surface followed by a top layer of 2 mg cm^{-2} of the optimal ionomer-Zr(OPr)$_4$ dispersion yielded similar initial polarization curves, but fast degradation of ~18 mV hour^{-1} (fig. S46) like that of ionomer-only systems (fig. S32). When ionomer is applied as a topcoat over Zr additives-ionomer that is in contact with the catalyst, it remains stable over 20 hours (figs. S48 and S49). This is evidence that the primary degradation mode is not chemical OH$^-$ attack but is associated with the oxidizing reaction environment. The oxidative degradation may be either direct or via a mediated process over a short length scale at the interface between catalyst and ionomer.

Although not in direct contact with electrocatalyst, the bulk membrane itself may degrade via nucleophilic substitution (e.g., attack by OH$^-$) and Hofmann-elimination reactions, leading to cation and ion-exchange capacity loss (*17*, *18*). Oxidative species generated during water oxidation (e.g., hydroxyl radicals, peroxides, and/or soluble high-valent metals) might also transport to the membrane from the catalyst-ionomer region and lead to degradation—for example, by backbone chain scission—although evidence for these processes in aryl-ether–free AEMs such as PiperION used here is inconclusive (*10*, *18*, *52*–*55*). Electron microscopy (fig. S50A) shows that the membrane retains its ~40-μm original thickness in inactive regions remote from the catalyst-ionomer–coated anode, but the sections in contact with the anode thinned and the membrane surface showed fiber-like indentations similiar to the geometry of the PTL (fig. S51). In AEMWEs without additive, the membrane thickness decreased by 20 and 44% after 20 and 100 hours at 2.0 A cm^{-2} and 70°C, respectively. The loss of membrane thickness was more apparent on the anode-facing side. Membrane thinning and etching were reduced in the AEMWEs with Zr(OPr)$_4$ additive, with 17 and 20% thinning after 20 and 100 hours, respectively, under identical conditions (fig. S50B).

Gel-permeation chromatography shows that the average molecular weight of the ionomer in the membrane from devices with and without Zr(OPr)$_4$ additive decreased after testing for 20 hours by 5.7 and 9.9% and after 100 hours by 9.7 and 16%, respectively (fig. S52). As the membrane is not in direct contact with the water-oxidation catalyst, the Zr additive also appears to slow remote oxidative degradation by decomposing or limiting transport of reactive species. To test this hypothesis, we treated ionomer with and without additive in 1 weight % (wt%) aqueous H_2O_2 at 70 °C for 12 hours. Nuclear magnetic resonance (NMR) analysis showed substantial oxidation of the aromatic region of the ionomer that was largely suppressed for the ionomer sample with Zr additive (fig. S9). Similarly, adding 100 parts

per million H_2O_2 to the anode pure-water feed in the optimal Zr additive-ionomer device had no apparent effect on durability (fig. S10).

Chemical interactions between additives and ionomer

Resonant small-angle x-ray scattering at energies near the Zr K-edge from films of ionomer with $Zr(OPr)_4$ additive (Fig. 4A) shows a prominent peak at scattering vector $q = 0.26$ Å^{-1} (d-spacing = 2.4 nm) that is absent in the ionomer-only sample. Nanoscale order appears driven by the interactions with Zr-oxo oligomers, water, and ionomer and is reminiscent of nanophase separation in perfluorosulfonic acid (Nafion) ionomers (*56*). This ordering may explain the increased ionic conductivity associated with channels that facilitate OH^- transport. The chemical interactions are evident in infrared spectra (Fig. 4B), which for the as-prepared ionomer have characteristic peaks at 1612 and 1379 cm^{-1} due to asymmetric stretching vibrations of C=O and C–O in HCO_3^- counterions (*57*, *58*). With $Zr(OPr)_4$, the intensity of these peaks decreases, whereas a new peak at 1563 cm^{-1} emerges and intensifies with increasing Zr content, which we attribute to coordination between Zr^{4+} and CO_3^{2-}. This assignment is supported by spectra of zirconium hydroxycarbonate [nominally $Zr(OH)_2CO_3{\cdot}ZrO_2$] with stretches at higher wave numbers than for Na_2CO_3 (fig. S56) owing to the stronger coordination bonds with higher-valent Zr^{4+} (*59*).

The chemical interactions between the additive species derived from $Zr(OPr)_4$ and the ionomer were further evident by dynamic light scattering (DLS). The PiperION ionomer dispersion in ethanol, diluted with isopropanol to match the concentration in the $Zr(OPr)_4$–ionomer solution, yielded particle-size distributions with peaks at 4 and 27 nm (Fig. 4C). In the solutions with $Zr(OPr)_4$ added to the ionomer, the primary peak shifts systematically to larger sizes as the ionomer:$Zr(OPr)_4$ ratio decreases, reaching ~100 nm for a ratio of 15:1. These results are consistent with infrared spectra that show that the Zr-additive species bind to carbonate counterions. The resulting negatively charged carbonate-capped Zr-oxo oligomers could act as inorganic chemical cross-links between the cationic ionomer backbones. To test this hypothesis, we prepared ionomer dispersions with Cl^- counterions. Minimal chemical interactions between Zr-additive species and ionomer(Cl^-) were found using the same spectroscopy and light-scattering analyses (fig. S58). With Cl^- counterions (fig. S59) and $Zr(OPr)_4$, the cell degradation rate was 3.8 ± 1.4 mV hour^{-1} over 20 hours at 2.0 A cm^{-2}, much higher than when (bi)carbonate was initially present in the additive-ionomer composite. Cryo-TEM analysis (fig. S60) revealed that the additives-ionomer(Cl^-) yielded Zr-containing nanoparticles resembling those in additives-ionomer(HCO_3^-), but with a less uniform distribution and aggregation into larger clusters. The precursor structure at the molecular level thus appears to be important for the formation of stable interphases with the anode catalyst. As another

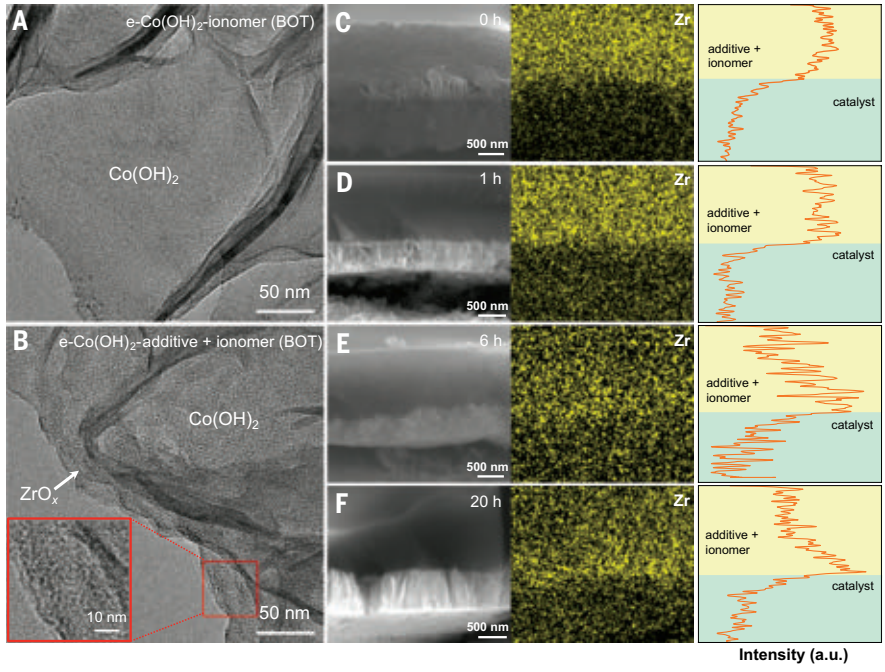


Fig. 3. Microscopic evidence of the protective interphase layer. Low-dose cryo-TEM of (**A**) e-$Co(OH)_2$-ionomer and (**B**) e-$Co(OH)_2$-additives-ionomer interphase before electrochemical testing. The inset in (B) is a magnified view of the outlined red area, highlighting well-dispersed nanometer-scale dark features consistent with ZrO_x dispersed in the ionomer. These samples were prepared by depositing e-$Co(OH)_2$ onto gold TEM grids, followed by coating of ionomers or additives-ionomer. (**C** to **F**) Cross-sectional SEM images and EDX maps of Zr with corresponding integrated intensity along the direction orthogonal to the catalyst-ionomer interface for (C) e-$Co(OH)_2$-additives-ionomer on PTLs and samples after testing at 2.0 A cm^{-2} for (D) 1 hour, (E) 6 hours, and (F) 20 hours.

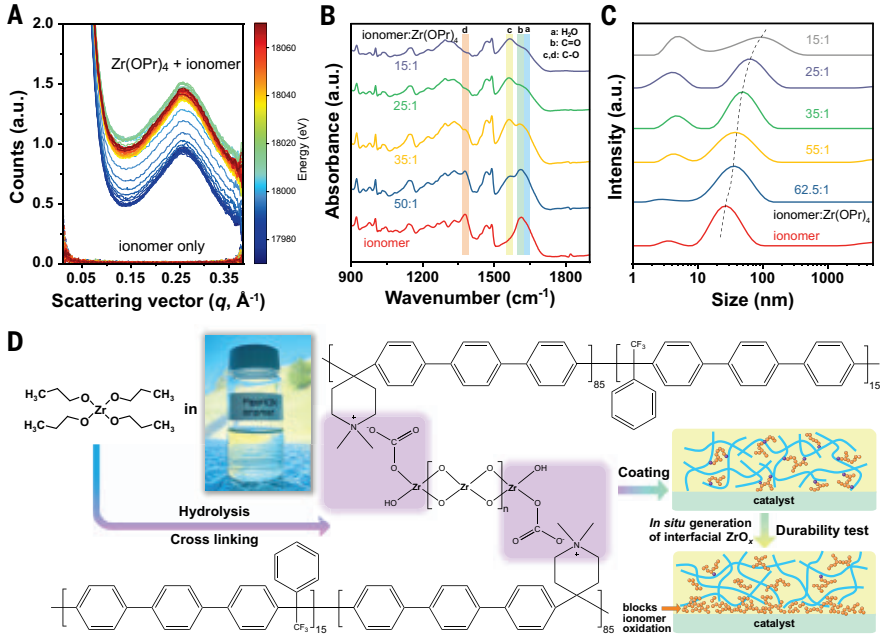


Fig. 4. Chemical interactions governing ionomer stabilization by inorganic additive. (**A**) Resonance small-angle x-ray scattering profiles of PiperION ionomer (bottom) and additives-ionomer (top) across the Zr K-edge. (**B**) Attenuated total reflectance Fourier-transform infrared (ATR-FTIR) spectra of ionomer and additives-ionomer thin films. (**C**) Dynamic light scattering analysis of ionomer dispersion and additive-ionomer sol solutions. (**D**) Schematic representation of the additives-ionomer preparation process and the underlying processes leading to protective interphase formation.

control, 20-nm ZrO_2 nanoparticles were mixed with ionomer dispersion and coated onto the catalyst surface, yielding a similarly high degradation rate of 7.9 mV hour^{-1} (fig. S65). Collectively, these findings highlight the probable role of molecular coordination bonds between the ionomer and Zr-additive species, here through bicarbonate or carbonate, that drive assembly of the inorganic-organic ionically conductive interphase, even though the (bi)carbonate is likely flushed from the system after extended operation.

In sum, we introduced and applied interphase engineering to ionomer-based electrochemical devices, yielding more than an order-of-magnitude improvement in durability for AEMWEs operating with pure-water feed. The molecular interactions between inorganic additive species formed via simple and scalable sol-gel chemistry and ionomers are central to creating the robust, ionically conductive material and interphase. During operation, the oxidatively inert ZrO_x inorganic component enriches at the catalyst interfaces and prevents continuous electrochemical oxidative degradation of the ionomer while reinforcing and preserving mechanical stability. Combining these advances with more-active electrocatalysts such as high-loading metallic perovskite oxides at the anode (*60*) and optimized PtRu catalysts at the cathode (*61*), alongside passivation layers (*62*) and iterative improvements to ionomer chemistry (*63*–*65*), may poise pure-water AEMWE as a disruptive commercial electrolysis technology for hydrogen production. More broadly, these approaches could be used to create interphases that control catalyst dissolution and redeposition via confinement (*66*), prevent oxidation side reactions in hydrocarbon-based ionomer devices such as PEM fuel cells (*67*), and tune interfacial adhesion and catalyst-layer mechanical properties.

REFERENCES AND NOTES

1. N. Du *et al.*, *Chem. Rev.* **122**, 11830–11895 (2022).
2. A. Buttler, H. Spliethoff, *Renew. Sustain. Energy Rev.* **82**, 2440–2454 (2018).
3. D. Li *et al.*, *Energy Environ. Sci.* **14**, 3393–3419 (2021).
4. A. Gladik, M. Riedel, R.-A. Eichel, *J. Power Sources* **628**, 235878 (2025).
5. D. Li *et al.*, *Nat. Energy* **5**, 378–385 (2020).
6. L. Wang *et al.*, *Energy Environ. Sci.* **10**, 2154–2167 (2017).
7. Y. Zheng *et al.*, *ACS Energy Lett.* **8**, 5018–5024 (2023).
8. M. Klingenhof *et al.*, *Nat. Catal.* **7**, 1213–1222 (2024).
9. J. R. Varcoe *et al.*, *Energy Environ. Sci.* **7**, 3135–3191 (2014).
10. E. J. Park *et al.*, *Chem. Soc. Rev.* **53**, 5704–5780 (2024).
11. S. W. Boettcher, *Chem. Rev.* **124**, 13095–13098 (2024).
12. P. Mardle, B. Chen, S. Holdcroft, *ACS Energy Lett.* **8**, 3330–3342 (2023).
13. R. A. Krivina *et al.*, *Adv. Mater.* **34**, e2203033 (2022).
14. G. Sakas *et al.*, *Int. J. Hydrogen Energy* **47**, 4328–4345 (2022).
15. S. Favero, I. E. L. Stephens, M. M. Titirci, *Adv. Mater.* **36**, e2308238 (2024).
16. N. Chen, Y. M. Lee, *Prog. Polym. Sci.* **113**, 101345 (2021).
17. M. G. Marino, K. D. Kreuer, *ChemSusChem* **8**, 513–523 (2015).
18. Q. Li *et al.*, *Chem. Sci.* **14**, 10429–10434 (2023).
19. J. Wang *et al.*, *Nat. Energy* **4**, 392–398 (2019).
20. W. Song *et al.*, *Nat. Commun.* **14**, 2732 (2023).
21. J. H. Hsu *et al.*, *Angew. Chem. Int. Ed.* **62**, e202304778 (2023).
22. W. You, K. J. T. Noonan, G. W. Coates, *Prog. Polym. Sci.* **100**, 101177 (2020).
23. C. H. Park *et al.*, *Nature* **532**, 480–483 (2016).
24. N. Chen *et al.*, *Nat. Commun.* **12**, 2367 (2021).
25. X. Hu *et al.*, *Nat. Energy* **9**, 401–410 (2024).
26. L. Wan *et al.*, *Energy Environ. Sci.* **16**, 1384–1430 (2023).
27. X. Wu, K. Scott, *Int. J. Hydrogen Energy* **38**, 3123–3129 (2013).
28. L. Xiao *et al.*, *Energy Environ. Sci.* **5**, 7869 (2012).
29. L. Wan *et al.*, *Small* **18**, e2200380 (2022).
30. J. Xiao *et al.*, *ACS Catal.* **11**, 264–270 (2020).
31. A. Li *et al.*, *Science* **384**, 666–670 (2024).
32. R. T. Liu *et al.*, *Chem. Soc. Rev.* **52**, 5652–5683 (2023).
33. G. A. Lindquist *et al.*, *Energy Environ. Sci.* **16**, 4373–4387 (2023).
34. J. Liu *et al.*, *J. Electrochem. Soc.* **168**, 054522 (2021).
35. N. U. Hassan, Y. Zheng, P. A. Kohl, W. E. Mustain, *J. Electrochem. Soc.* **169**, 044526 (2022).
36. B. Mayerhöfer *et al.*, *Int. J. Hydrogen Energy* **47**, 4304–4314 (2022).
37. J. Lim *et al.*, *ACS Energy Lett.* **9**, 3074–3083 (2024).
38. D. Li *et al.*, *ACS Appl. Mater. Interfaces* **11**, 9696–9701 (2019).
39. E. K. Volk *et al.*, *ACS Electrochem.* **1**, 239–248 (2024).
40. R. A. Krivina *et al.*, *ACS Appl. Mater. Interfaces* **14**, 18261–18274 (2022).
41. J. B. Goodenough, K. S. Park, *J. Am. Chem. Soc.* **135**, 1167–1176 (2013).
42. Y. S. Meng, V. Srinivasan, K. Xu, *Science* **378**, eabq3750 (2022).
43. X. Fan *et al.*, *Sci. Adv.* **4**, eaau9245 (2018).
44. Y. Gao *et al.*, *Nat. Mater.* **18**, 384–389 (2019).
45. B. Jagger, M. Pasta, *Joule* **7**, 2228–2244 (2023).
46. C. J. Brinker, G. W. Scherer, *Sol-gel science: The physics and chemistry of sol-gel processing* (Academic press, 2013).
47. A. Veyland, L. Dupont, J. Rimbault, J.-C. Pierrard, M. Aplincourt, *Helv. Chim. Acta* **83**, 414–427 (2000).
48. K. Li, J. G. Chen, *ACS Catal.* **9**, 7840–7861 (2019).
49. L. Chen, Q. Xu, S. W. Boettcher, *Joule* **7**, 1867–1886 (2023).
50. N. Ziv, D. R. Dekel, *Electrochem. Commun.* **88**, 109–113 (2018).
51. M. F. Ernst, V. Meier, M. Kornherr, H. A. Gasteiger, *J. Electrochem. Soc.* **171**, 074511 (2024).
52. L. Ghassemzadeh, S. Holdcroft, *J. Am. Chem. Soc.* **135**, 8181–8184 (2013).
53. A. Moysiadou, S. Lee, C. S. Hsu, H. M. Chen, X. Hu, *J. Am. Chem. Soc.* **142**, 11901–11914 (2020).
54. L. Xu, K. H. Møller, J. D. Crounse, H. G. Kjaergaard, P. O. Wennberg, *Environ. Sci. Technol.* **54**, 13467–13477 (2020).
55. X. Zhang *et al.*, *Environ. Sci. Technol.* **55**, 623–633 (2021).
56. G. M. Su *et al.*, *J. Am. Chem. Soc.* **141**, 13547–13561 (2019).
57. N. J. Firet, W. A. Smith, *ACS Catal.* **7**, 606–612 (2016).
58. M. Moradzaman, G. Mul, *ACS Catal.* **10**, 8049–8057 (2020).
59. K. D. Dobson, A. J. McQuillan, *Langmuir* **13**, 3392–3396 (1997).
60. T. Zhai *et al.*, *J. Am. Chem. Soc.* **147**, 15448–15458 (2025).
61. A. Shirwalkar *et al.*, *ACS Energy Lett.* **10**, 1779–1785 (2025).
62. M. Kwak, K. Ojha, M. Shen, S. W. Boettcher, *ACS Energy Lett.* **9**, 1025–1034 (2024).
63. X. Wu, N. Chen, H. A. Klok, Y. M. Lee, X. Hu, *Angew. Chem. Int. Ed.* **61**, e202114892 (2022).
64. C. Hu *et al.*, *Adv. Sci.* **11**, e2306988 (2024).
65. L. Yin *et al.*, *Angew. Chem. Int. Ed.* **63**, e202400764 (2024).
66. A. D. Pendergast, S. W. Boettcher, *Science* **387**, 718–719 (2025).
67. Z. Han *et al.*, *ACS Electrochem.* **1**, 588–598 (2024).
68. S. Hou *et al.*, Dataset of "Durable, pure-water-fed, anion-exchange-membrane electrolyzers through interphase engineering," Zenodo (2025); https://doi.org/10.5281/zenodo.14868077

ACKNOWLEDGMENTS

We thank G. Freychet for assisting with the collection of x-ray scattering data at Brookhaven National Laboratory. R.T.H. gratefully acknowledges D. H. Marin for insightful discussions regarding the online ICP-MS measurements. **Funding:** The electrolyzer device work was supported by the US Department of Energy (DOE), Office of Energy Efficiency and Renewable Energy, award DE-EE0011322. The fundamental work of understanding additive-ionomer and catalyst interface interactions was funded by the DOE Center for Ionomer-Based Water Electrolysis, award DE-AC02-05CH11231. The work made use of the Molecular Foundry, a user facility supported by the DOE under contract DE-AC02-05CH11231. R.T.H., V.C., and T.F.J. acknowledge funding from the Office of Naval Research, grant N00014-24-1-2433, for the design of the online ICP-MS system. Partial support for the online ICP-MS measurements came from the US DOE, Office of Science, Office of Basic Energy Sciences, Chemical Sciences, Geosciences, and Biosciences Division, Catalysis Science Program to the SUNCAT Center for Interface Science and Catalysis. The cryo-TEM and EELS portions of this research were supported by the Center for Nanophase Materials Sciences (CNMS), which is a DOE User Facility at Oak Ridge National Laboratory. Development of x-ray scattering protocols for ionomer composites was supported by the DOE under contract no. DE-AC02-05CH11231 within the Inorganic/Organic Nanocomposites Program (KC3104). K.K.-Y.L. and G.M.S. acknowledge support from the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory, under DOE contract DE-AC02-05CH11231. This research used the Soft Matter Interfaces beamline at the National Synchrotron Light Source II, a DOE user facility operated under contract DE-SC0012704, and beamline 7.3.3 of the Advanced Light Source, which is a DOE Office of Science User Facility under contract DE-AC02-05CH11231. **Author contributions:** S.H. conceived and designed the experiments and completed most of the synthesis, characterization, and electrolyzer tests. S.W.B. proposed and supervised the entire project. S.H., Y.Z., and S.W.B. wrote the paper with input from all coauthors. A.S. and B.L. performed the scanning electron microscope (SEM) cross-section, GPC, and ion-exchange capacity for membrane degradation. J.O. and M.J.Z. conducted the cryo-TEM and EELS analyses. M.K. performed cross-sectional electron microscopy and x-ray photoelectron spectroscopy analysis. K.K.-Y.L. and G.M.S. performed x-ray scattering measurements and analysis. P.W. and J.F. contributed to the DLS analysis. R.T.H., V.C., and T.F.J. conducted the online ICP-MS experiments. A.C.E. performed the NMR measurements. H.D. and Y.Y. contributed to the hydroxide conductivity and water uptake experiments. All authors discussed the results and provided editing and comments on the manuscript. **Competing interests:** Patent applications were filed on the additive technology (US Provisional Application no. 63/291,295) with some of the coauthors as inventors. A.S. and B.L. are employees of Versogen and Y.Y. is chief executive officer and founder of Versogen, a company working to commercialize AEMWE technology that makes and markets PiperION membranes and ionomers used in this study. The authors declare that they have no other conflicts of interest. **Data and materials availability:** All data are available in the main text or supplementary materials. Polarization curves, chronopotentiometry stability data of AEMWE cells, cryo-TEM and SEM images, energy-dispersive x-ray (EDX) maps, ATR-FTIR spectra, and DLS data in the main text are available at Zenodo (*68*). **License information:** Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adw7100
Materials and Methods; Figs. S1 to S72; Table S1; References (*69*–*99*)

Submitted 15 February 2025; accepted 21 August 2025

10.1126/science.adw7100

Electrically controlled interlayer trion fluid in electron-hole bilayers

Ruishi Qi[1,2], Qize Li[1], Zuocheng Zhang[1]†, Sudi Chen[1,2], Jingxu Xie[1,2], Yunbo Ou[3], Zhiyuan Cui[1,2], David D. Dai[4], Andrew Y. Joe[5], Takashi Taniguchi[6], Kenji Watanabe[7], Sefaattin Tongay[3], Alex Zettl[1,2,8], Liang Fu[4], Feng Wang[1,2,8]*

The combination of repulsive and attractive Coulomb interactions in a quantum electron-hole (e-h) fluid can produce correlated phases of multiparticle charge complexes, such as excitons, trions, and biexcitons. We report an experimental realization of an electrically controlled interlayer trion fluid in van der Waals heterostructures. In strongly coupled e-h bilayers, electrons and holes spontaneously form three-particle trion bound states. The interlayer trions can assume 1e-2h and 2e-1h configurations. We show that the two holes in 1e-2h trions form a spin-singlet with a spin gap of approximately one milli–electron volt. By electrostatic gating, the equilibrium state can be continuously tuned into an exciton fluid, a trion fluid, an exciton-trion mixture, or a trion-charge mixture. Our work demonstrates a platform to study correlated phases of tunable Bose-Fermi mixtures.

In many-body systems composed of oppositely charged particles, the combination of repulsive and attractive Coulomb interactions can stabilize a wide variety of multiparticle bound states. For example, interacting electrons and nuclei form atoms and molecules, and interacting electrons and positrons generate two-particle positronium and three-particle positronium ions. In semiconductors, electrons and holes can form excitons, trions, and biexcitons (*1*). These multiparticle electron-hole (e-h) states in semiconductors are usually generated transiently through optical excitation, which makes it difficult to explore their steady-state quantum phases because of their short lifetimes (*2*–*4*). van der Waals heterostructures composed of two transition metal dichalcogenide (TMD) semiconductor layers separated by a thin hexagonal boron nitride (hBN) layer enable the creation of electrically controlled e-h bilayers. This configuration opens opportunities to explore multiparticle e-h complexes at thermal equilibrium (*5*–*11*). Previously, electrically controlled interlayer excitons (i.e., correlated e-h bound pairs) have been realized in $MoSe_2$/hBN/WSe_2 heterostructures, leading to the observation of interlayer excitonic insulators (EIs) (*6*, *7*) and perfect Coulomb drag behavior (*12*, *13*). Such exciton fluids are expected to host high-temperature exciton condensates (*8*, *14*). Three- and four-particle e-h complexes have also been theoretically predicted to exist in e-h bilayers in the strong coupling limit, potentially leading to distinct types of quantum behaviors (*10*, *15*–*18*). Their experimental realization, however, has remained elusive.

We report an experimental observation of a stable quantum fluid of interlayer trions—three-particle e-h complexes—in an e-h bilayer. We demonstrate that when strongly coupled e-h fluids are electrostatically tuned to a commensurate e-h density ratio of 2:1 or 1:2, Coulomb attraction together with quantum effects leads to the spontaneous creation of negative or positive trions (*15*)—two electrons bound to one hole or two holes bound to one electron—in thermal equilibrium. Direct evidence of trion formation is revealed by optical spectroscopy.

[1]Department of Physics, University of California, Berkeley, CA, USA. [2]Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [3]School for Engineering of Matter, Transport and Energy, Arizona State University, Tempe, AZ, USA. [4]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA, USA. [5]Department of Physics, University of California, Riverside, CA, USA. [6]Research Center for Materials Nanoarchitectonics, National Institute for Materials Science, Tsukuba, Japan. [7]Research Center for Electronic and Optical Materials, National Institute for Materials Science, Tsukuba, Japan. [8]Kavli Energy NanoSciences Institute, University of California Berkeley, and Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ***Corresponding author. Email: fengwang76@berkeley.edu** †Present address: Department of Physics and Astronomy, University of Nebraska–Lincoln, Lincoln, NE, USA.

Strongly coupled e-h fluids

Figure 1A schematically shows the e-h bilayer device. A monolayer MoS_2 (*19*, *20*) and a monolayer WSe_2 are separated by a monolayer or bilayer hBN tunneling barrier. Owing to the type-II band alignment, the electrons and holes in the system naturally reside in the MoS_2 and WSe_2 layers, respectively. Such an e-h bilayer structure provides a highly tunable platform for controlling and studying a strongly correlated Bose-Fermi mixture. Applying gate voltages and interlayer bias voltage tunes the band alignment and the Fermi energy of the two layers (*10*, *21*), enabling the observation of a long-lived quasi-equilibrium dipolar EI phase at zero magnetic field (*6*, *7*, *12*, *13*). When the EI is doped with additional electrons or holes, the interlayer excitons can, in principle, bind with additional unpaired charges to create interlayer trions. However, it is more challenging to achieve spontaneous formation of interlayer trions in thermal equilibrium because they only appear in the very strong e-h coupling limit owing to their small binding energy (*22*). In this work, we fabricate e-h bilayer devices with substantially reduced interlayer distance (within 1 nm) compared with those from previous studies (*6*, *7*, *12*, *13*). The enhanced e-h attraction relative to the electron-electron and hole-hole repulsion favors trion formation. With large e-h momentum mismatch suppressing the interlayer recombination and specially designed carrier reservoir ensuring low contact resistance [fig. S1 (*23*)], the e-h fluids remain near-equilibrium with the electrodes. The heterostructure is encapsulated by dielectric hBN on both sides and gated by few-layer graphene top gate (TG) and bottom gate (BG). We keep the electron layer grounded ($V_e = 0$) and apply voltages on the gates and the hole layer (V_h). The symmetric gate voltage $V_G \equiv V_{TG} + V_{BG}$ tunes the Fermi-level and thus the net charge density (e-h imbalance). With the antisymmetric gating $V_{TG} - V_{BG}$ fixed throughout our experiment, the bias voltage $V_B \equiv V_h - V_e$ tunes the e-h pair density. Figure 1B is an optical image of such a device (D1) with a monolayer hBN spacer. Although we focus on this device in the main text, our conclusions are also supported by highly reproducible data taken from another three devices (D2, D3, and D4), as summarized in figs. S2 to S4 (*23*). The samples are kept at a temperature $T = 4.9$ K unless otherwise specified.

The optical absorption spectrum of TMDs is sensitive to the charge density because of strong excitonic resonances and Coulomb interactions (*24*, *25*). Taking advantage of this dependence, we determine the electron and hole densities as a function of applied voltages using the spectroscopy technique described by Qi *et al.* (*7*). Figure 1C shows the charge doping phase diagram as a function of V_G and V_B, where red and green channels of the false color map encode the density of electrons (n_e) and the density of holes (n_h), respectively. Black dashed lines approximately trace the commensurate density ratios between n_e and n_h. Figure 1, D and E, shows two representative gate-dependent reflectivity scans at constant bias voltages $V_B = 0.68$ V and 0.735 V, respectively. The energy-derivative of reflectivity, dR/dE, is plotted for better visibility. At low bias voltages, such as the one in Fig. 1D, the system has a finite type-II bandgap, so only one type of charge can enter the system at a time. Well-defined intralayer exciton peaks (X_0) for MoS_2 and WSe_2 are observed when both layers are undoped. Upon electron or hole doping, another absorption peak emerges at lower energy on the MoS_2 or WSe_2 side, which is commonly known as the intralayer trion peak or the charged exciton peak (X^-/X^+) (*24*, *25*). When the bias voltage exceeds the gap energy at $V_B > 0.72$ V, the equilibrium state of the system contains a fluid of electrons and holes at the same time, as signified by the coexistence of the WSe_2 X^+ peak and MoS_2 X^- peak

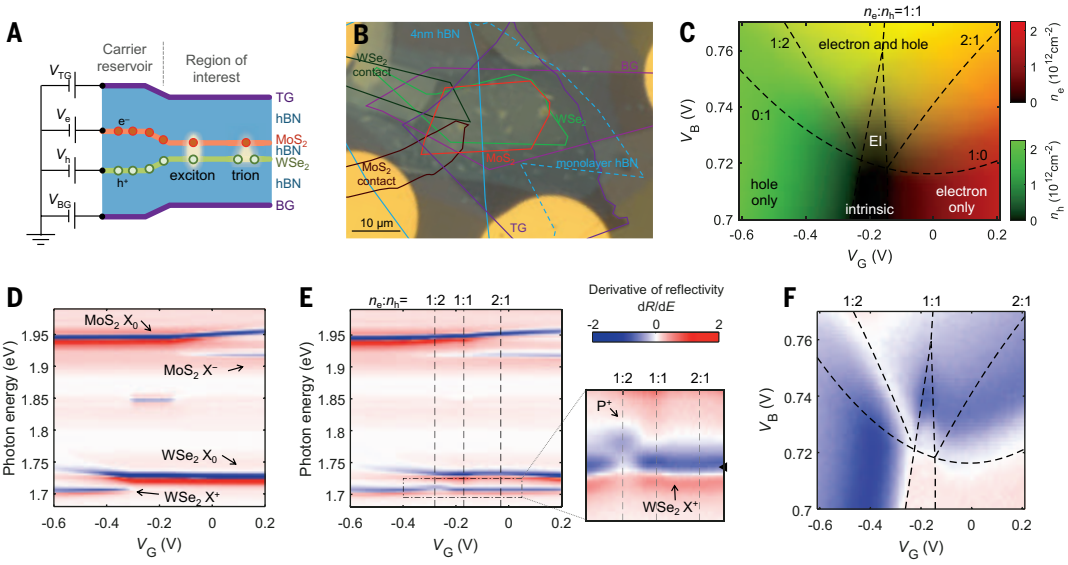


Fig. 1. Electrically controlled e-h fluids. (**A**) Schematic cross section of the MoS_2/hBN/WSe_2 heterostructure device. (**B**) An optical image of the MoS_2/monolayer hBN/WSe_2 heterostructure device D1, with flake boundaries outlined. (**C**) Carrier doping phase diagram of the e-h bilayer. The vertical electric field is fixed by $V_{TG} - V_{BG} = 5.6$ V, whereas the gate voltage V_G and bias voltage V_B are varied. The red and green channels of the image show the electron density and the hole density, respectively. (**D** and **E**) Energy-derivative of reflection spectrum dR/dE as a function of gate voltage V_G. The bias voltage is fixed at $V_B = 0.68$ V (D) and $V_B = 0.735$ V (E). (**F**) dR/dE at WSe_2 X^+ resonance energy [1.707 eV, marked by the black triangle in the inset of (E)]. (D), (E), and (F) share the same color scale.

in Fig. 1E. In this regime, strong Coulomb attraction between the electrons and holes can result in spontaneous formation of correlated bound states in thermal equilibrium. The e-h fluids can possibly consist of interlayer excitons, interlayer trions, unpaired electrons and/or holes, or a mixture of them.

We pay special attention to three states with commensurate density ratios n_e:n_h = 1:1, 2:1, and 1:2. At n_e:n_h = 1:1, electrostatically injected electrons and holes are known to pair into interlayer excitons and form a strongly correlated EI (*6*, *7*). The formation of the interlayer excitons has only a weak effect on the optical spectra: Careful examination shows that the WSe_2 X^+ peak has a slight red shift of ~0.2 meV (fig. S5). At n_e:n_h = 1:2, we note that the WSe_2 X^+ resonance exhibits a prominent feature (Fig. 1E). Compared with nearby regions, the WSe_2 X^+ peak almost disappears at this density ratio, and a new peak (labeled as P^+) emerges at slightly higher energy. A similar feature, although less obvious, is observed on the MoS_2 side at the carrier ratio n_e:n_h = 2:1 (fig. S6). Although the spectral features are relatively weak, they are very reliable: The spectral change is much stronger than experimental uncertainties and is highly reproducible over multiple samples. The emerging peak centered at n_e:n_h = 1:2 and 2:1 is an early indication of the formation of positive and negative interlayer trions, respectively. Because the WSe_2 X^+ peak is sharper and stronger, we will focus our analysis on the WSe_2 resonances. Figure 1F shows the dR/dE signal at the WSe_2 X^+ peak energy over the entire e-h density phase space. We can observe a clear feature associated with the decreased signal at the WSe_2 X^+ resonance along the 1:2 ratio line.

Spectroscopic signature of trions

Next, we confirm that the P^+ peak emerges thanks to the formation of interlayer trions. It is well known that hydrogen anions—two electrons bound to one proton—have only one bound state, with two electrons forming a spin-singlet. Trions are correlated three-particle e-h complexes that resemble hydrogen anions. A definitive signature of the trion is its internal spin structure: The two same-charge particles must form a spin-singlet pair to lower the energy (*15*, *26*). The magnetic field dependence of the optical transitions can therefore provide an unambiguous spectroscopic signature of the interlayer trions.

Spin-valley locking combined with the optical selection rules in semiconducting TMDs allow us to selectively access the holes of a particular spin using a circularly polarized laser probe (*27*–*29*). Figure 2, A and B, compares the out-of-plane magnetic field (B) dependence of helicity-resolved dR/dE spectra at a fixed hole density, $n_h = 1.0 \times 10^{12}$ cm^{-2}, but different electron doping, $n_e = 0$ and $n_e = n_h/2$. The spectra are taken with left-handed circularly polarized light that selectively probes the spin-up component of the holes. When $n_e = 0$, the holes form a Fermi liquid, and the spin can be continuously polarized by the external magnetic field. The X^+ resonance intensity therefore shows a monotonic increase with the applied magnetic field.

By contrast, distinct field dependence is observed when $n_e = n_h/2$ (Fig. 2B). We observe two peaks with an energy separation of ~5 meV, and they exhibit completely different magnetic field dependence. The X^+ peak (marked by the yellow dashed line) still shows an increase in intensity with the magnetic field. However, its intensity is much lower than in the $n_e = 0$ case. The new peak P^+ (marked by the purple dashed line) does not change monotonically with B; instead, it is the strongest at zero field and decays symmetrically at higher field regardless of the field direction (Fig. 2B, inset). The unusual magnetic field dependence of the P^+ arises from the spin-singlet configuration of the two holes in the trion: The singlet cannot be continuously polarized. Instead, spin-singlet trions dissociate at high field when the Zeeman energy becomes comparable to the trion binding energy.

The resonant photon absorption process at zero magnetic field is schematically illustrated in Fig. 2C. When only free holes are initially present ($n_e = 0$), the intralayer trion resonance (X^+ peak) can be excited when the free hole and the photoexcited hole have opposite spin (and valley). At the density ratio $n_e = n_h/2$, the system forms positive interlayer trions in thermal equilibrium, where there are already two identical holes with singlet pairing bound to one electron. The optically generated e-h pair will interact with them in a distinctly different way. The five-particle resonant state in this case involves three identical holes in the WSe_2 layer, two of which must have the same spin, increasing the total energy because of the Pauli exclusion principle. It can be understood in analogy to the instability of negative molecular hydrogen ions H_2^- (*30*), where three electrons cannot all occupy the

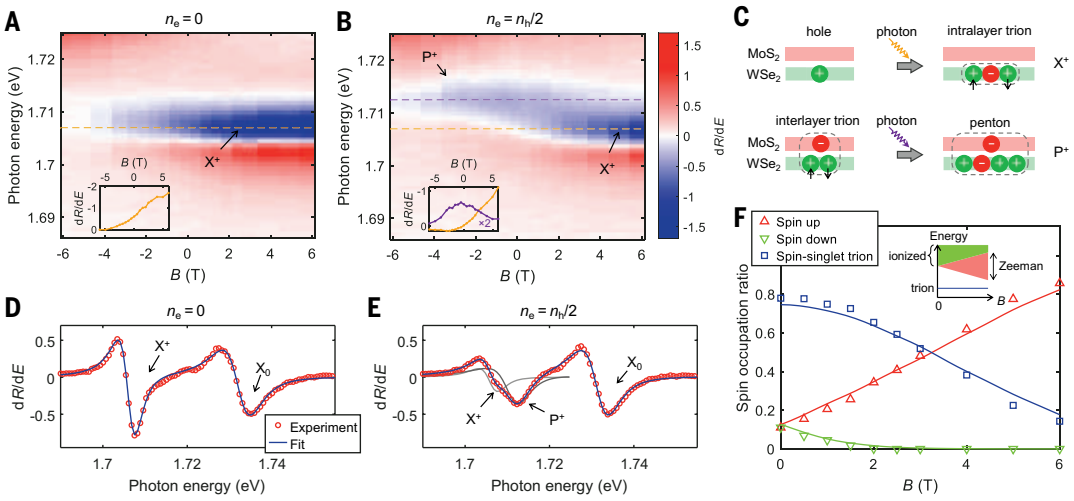


Fig. 2. Spectroscopic evidence of interlayer trion formation. (**A**) dR/dE spectrum at varying perpendicular magnetic field B, with left-handed circularly polarized light probe. Carrier densities are $n_h = 1.0 \times 10^{12}$ cm^{-2} and $n_e = 0$. (Inset) dR/dE at X$^+$ peak energy as a function of B. (**B**) As in (A), but with $n_h = 1.0 \times 10^{12}$ cm^{-2} and $n_e = 0.5 \times 10^{12}$ cm^{-2}. (Inset) dR/dE at X$^+$ (yellow) and P$^+$ (purple) peak energy as a function of B. The purple line is multiplied by 2 for visibility. (**C**) Schematic illustration of resonant photon absorption process for two representative doping ratios. Red and green circles are electrons and holes, respectively. Dashed gray boxes denote bound states. Black arrows represent spins. (**D** and **E**) Comparison of the dR/dE spectrum between $n_e = 0$ (D) and $n_e = n_h/2$ (E) at zero magnetic field. The experimental spectra (open circles) are fitted to the derivative of multiple Fano peaks (solid lines). (**F**) Magnetic field dependence of spin-up, spin-down, and spin-singlet trion state occupation ratio, at $n_h = 1.0 \times 10^{12}$ cm^{-2} and $n_e = 0.5 \times 10^{12}$ cm^{-2}. Symbols are experimental data from (B), and solid lines are fitting results from the model described in the supplementary materials (*23*). (Inset) Illustration of the B-induced energy shift for the singlet trion state (blue) and the continuum of ionized states (red and green for two spins).

lowest molecular orbital. The energy cost for occupying a higher orbital leads to the P$^+$ peak at higher energy.

Figure 2, D and E, compares the zero-field optical spectra at $n_e = 0$ and $n_e = n_h/2$. At $n_e = 0$, the system contains a well-understood two-dimensional (2D) hole gas in WSe_2, where two absorption peaks, the intralayer exciton peak X_0 and the intralayer trion peak X$^+$, are observed. Each of them can be well described by a Fano line shape. At $n_e = n_h/2$, the spectrum can be described by three Fano peaks. The X$^+$ peak loses most of its oscillator strength and the high-energy P$^+$ peak appears. The X$^+$ peak does not completely disappear at $n_e = n_h/2$, which suggests that a portion of the interlayer trions are ionized owing to the finite temperature and small trion binding energy.

The magnetic field dependence of optical spectra allows us to estimate the interlayer trion binding energy ε_t. At zero magnetic field, the ground state of the system is the spin-singlet trion, and excited states include a continuum of ionized states (unbound hole-exciton pairs) that are spin degenerate. When the field is weak, most holes are in the trion bound state, alongside a small portion of thermally ionized holes. With a strong external field, the Zeeman effect splits the energy levels of the spin-up and spin-down ionized states, but the trion state energy is not affected. This reduces the gap between the ground state and the continuum, ionizing a larger portion of the trions (Fig. 2F, inset). Assuming that the oscillator strength of each optical transition peak depends linearly on the density of the corresponding particle, the spin-up and spin-down ionized hole density will be proportional to the X$^+$ oscillator strength at B and $-B$, respectively, and the density of holes in the trion bound state will be proportional to the P$^+$ peak strength. By fitting the oscillator strength of each peak, we get the ratio of the occupation number of each spin state as a function of B (Fig. 2F). With an effective hole g factor of $g = 6.1$ (*31*), a 1-T field amounts to 0.35-meV Zeeman shift. The field required to ionize half of the trions is roughly 4 T, indicating a trion binding energy on the order of 1 meV. We fit the experimental data to a simple model (*23*) and estimate the spin gap, or the trion binding energy, to be $\varepsilon_t = 1.1 \pm 0.3$ meV. It is one to two orders of magnitude smaller than the interlayer exciton binding energy ($\varepsilon_x = 42 \pm 5$ meV; fig. S7) (*6*, *7*). This binding energy ratio is similar to the hydrogen anion compared with the hydrogen atom (0.75 eV versus 13.6 eV) and is also consistent with previous quantum Monte Carlo simulation results (*22*). The temperature dependence of the trion occupation shown in fig. S8 (*23*) also suggests a binding energy of ~1 meV, where half of the trions ionize at $T \approx 10$ K.

Electrically tunable Bose-Fermi mixtures

Because we can control the e-h density ratio by electrostatic gating, the ground state of the system can be electrically tuned to form different mixtures of excitons, trions, and unpaired charges. Figure 3A shows the WSe_2 resonances as a function of n_e for a fixed hole doping density $n_h = 1.0 \times 10^{12}$ cm^{-2}. The spectrum remains almost the same for $n_e > n_h$ as well as $n_e = 0$ but shows an evident change between $0 < n_e < n_h$ centered at $n_e = n_h/2$. Figure 3B shows the fitted oscillator strength of the X$^+$ peak and the P$^+$ peak as a function of n_e. Starting from the 2D hole gas at $n_e = 0$, adding electrons into the system creates interlayer trions, decreasing the number of free holes. This leads to an initial drop in X$^+$ intensity and an increase in P$^+$ intensity. Many-body exciton polaron interactions could be present in the trion-hole mixture. The P$^+$ peak energy does not shift substantially with n_e here, which indicates that the polaron effect is relatively weak at the low hole doping regime probed in this study. The density ratio $n_e = n_h/2$ is most favorable for interlayer trions. The X$^+$ oscillator strength decreases to ~20% at this ratio, indicating that ~80% of the holes are in the positive trion state at our base temperature $T = 4.9$ K, which amounts to a positive trion density $n_t^+ \approx 0.4 \times 10^{12}$ cm^{-2}. When n_e is further increased, the additional electrons will dissociate some interlayer trions because the exciton binding energy is larger than the trion binding energy, leading to a mixture of interlayer excitons and positive trions. At net charge neutrality $n_e = n_h$, the system turns into an EI phase (fig. S7). Further increasing the electron density makes negative trions favorable. Negative trions are composed of two electrons but only one hole, so they do not alter the WSe_2 absorption spectrum considerably because binding with the photoexcited WSe_2 e-h pair does not involve three identical particles. Signatures of negative trions are manifested in a high-energy peak (P$^-$) in the MoS_2 spectrum (fig. S6).

A 2D map of the positive trion density, experimentally estimated using the P^+ peak intensity, is displayed as a function of n_e and n_h in Fig. 3C. We consider a simple model at zero temperature to explain the findings. Starting from n_e electrons and n_h holes, they will first pair into excitons to lower the total energy. The exciton density would be $\min(n_e,n_h)$, limited by the minority carrier. Then the $\min(n_e,n_h)$ excitons and $|n_e-n_h|$ unpaired charges will further bind into a maximum number of trions, which is again limited by the smallest of the three concentrations. Thus, the trion density in this simple picture is given by $n_t = \min(n_e, n_h, |n_e-n_h|)$. There will be $n_x = \min(n_e,n_h) - n_t$ interlayer excitons and $n_u = |n_e-n_h| - n_t$ unpaired charges. Figure 3D shows the interlayer trion (including positive and negative trions) density as a function of n_e and n_h according to this model. The trion density is zero along the diagonal line (interlayer exciton fluid), the horizontal axis (2D electron gas), and the vertical axis (2D hole gas). The two peaks along the n_e:n_h = 1:2 line and the 2:1 line correspond to positive and negative trions, respectively. The positive trion sector agrees nicely with the experimental estimation using the WSe_2 spectra, whereas the negative trion sector is also similar to the MoS_2 results shown in fig. S6 (*23*). Apart from these commensurate ratios, the ground state of the e-h bilayer can be continuously tuned to be a hole-trion mixture ($n_e < n_h/2$), an exciton-trion mixture ($n_h/2 < n_e < 2n_h$), or an electron-trion mixture ($n_e > 2n_h$), as illustrated in Fig. 3E.

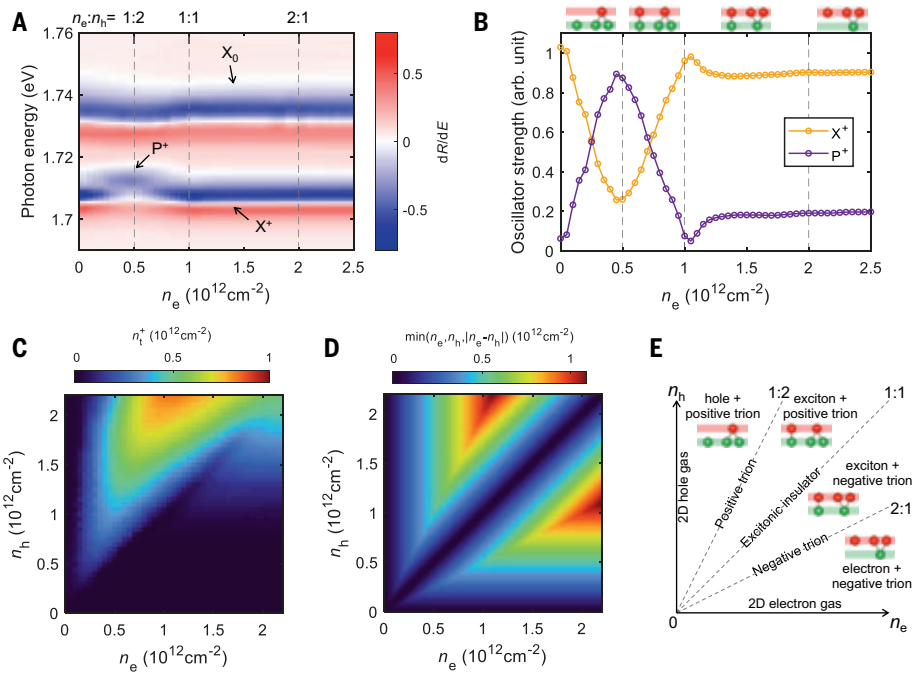


Fig. 3. Tunable trion-exciton-charge mixture. (**A**) dR/dE spectrum as a function of n_e for fixed $n_h = 1.0 \times 10^{12}$ cm^{-2}. (**B**) Fitted X^+ and P^+ oscillator strengths as a function of n_e for fixed $n_h = 1.0 \times 10^{12}$ cm^{-2}. (**C**) Experimentally estimated positive trion density n_t^+ as a function of n_e and n_h. (**D**) Trion density from the simple physical picture described in the text. (**E**) Schematic phase diagram of the e-h fluid.

High-order multiparticle complexes

The exciton fluid and the trion fluid can also host higher-order multiparticle states upon photoexcitation. We have already seen the prominent P^+ peak at $n_e = n_h/2$ as an example, corresponding to five-particle complexes created by photoexcited intralayer e-h pairs interacting with positive interlayer trions. For the case of $n_e = 0$, $n_e = n_h$, and $n_e = 2n_h$, we only observe a X^+ peak, and the reflection spectra are very similar among them. However, there is a small spectral difference at these density ratios (Fig. 4A). Their resonance energy, together with the interlayer exciton binding energy $\varepsilon_x = 42 \pm 5$ meV and interlayer trion binding energy $\varepsilon_t = 1.1 \pm 0.3$ meV, allows us to determine the relative energy levels of three-particle, four-particle, and five-particle bound states (Fig. 4, B to F).

We start with the simple three-particle case, where the photoexcited e-h pair interacts with the electrostatically injected hole. The X^+ peak can be well fitted by a Fano resonance (Fig. 4A, blue curve). Subtracting the X_0 peak energy in the low doping limit (1.727 eV) from the fitted energy gives the intralayer trion binding energy of $\varepsilon_3 = 20.6$ meV (Fig. 4C).

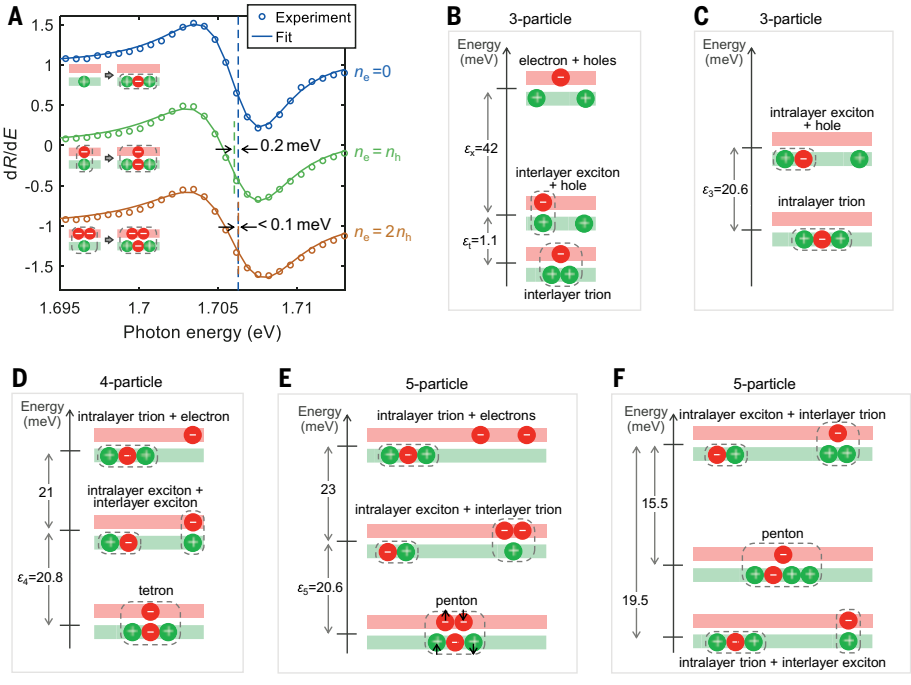


Fig. 4. Photoexcited high-order multiparticle charge complexes. (**A**) Comparison of the X^+ peak for $n_e = 0$, n_h, and $2n_h$, keeping $n_h = 1.0 \times 10^{12}$ cm^{-2}. Vertical dashed lines mark the fitted center energy of the Fano peak. (Inset) Illustrations of the corresponding photon absorption process. (**B** to **F**) Energy diagrams of multiparticle complexes with one MoS_2 electron and two WSe_2 holes (B); one WSe_2 electron and two WSe_2 holes (C); one MoS_2 electron, one WSe_2 electron, and two WSe_2 holes (D); two MoS_2 electrons, one WSe_2 electron, and two WSe_2 holes (E); and one MoS_2 electron, one WSe_2 electron, and three WSe_2 holes (F).

Next, we consider the X^+ peak at $n_e = n_h$ (EI phase). Comparing the X^+ spectrum (Fig. 4A, green curve) with the $n_e = 0$ case, there is a very small energy shift of 0.2 ± 0.1 meV. The energy shift is reproducible among multiple devices and follows a systematic trend when the interlayer distance is varied, reaching a maximum value of 0.4 ± 0.1 meV for bilayer hBN spacer devices (fig. S5). It suggests the formation of optically generated tetron state (also referred to as a Bose polaron as opposed to the Fermi polaron) (*32*) when the system has an EI equilibrium state. As schematically illustrated in Fig. 4D, the tetron can be viewed as an interlayer-intralayer hybrid biexciton state. The fitted peak energy is $\varepsilon_4 = 20.8$ meV lower than the X_0 peak, indicating a slightly higher binding energy than the intralayer trion. The energy difference being small and the interlayer distance dependence are both qualitatively consistent with a recent theoretical prediction (*32*). Despite the small spectral change, they correspond to distinct objects.

Comparison between the resonance energy in the 2D hole gas ($n_e = 0$; Fig. 4A, blue curve) and in the negative trion fluid ($n_e = 2n_h$; Fig. 4A, brown curve) gives a vanishingly small energy difference. It indicates that the five-particle bound state (two MoS_2 electrons, one WSe_2 electron, and two WSe_2 holes; illustrated in Fig. 4E) has almost the same binding energy of 20.6 meV as the WSe_2 intralayer trion. This behavior is very different from the high-energy P^+ peak discussed earlier in the positive trion case. We note that the five-particle state in this case does not involve three identical particles: Two MoS_2 electrons and two WSe_2 holes can both form spin-singlets and do not cost additional energy owing to aligned spins. Compared with the WSe_2 intralayer trion state, the absorption resonance energy is minimally affected by the presence of electrons in the opposite layer.

Finally, Fig. 4F illustrates the excited states of the positive trion fluid after photoexcitation, or the P^+ peak discussed earlier. The binding energy of this penton—an electron in MoS_2, an electron in WSe_2, and three holes in WSe_2 all bound together—can be determined from the fitting in Fig. 2E. It is 15.5 meV lower than the energy of an interlayer trion far separated from an intralayer exciton. However, this is not the lowest-energy state of these five particles. Because the interlayer trion binding energy is very small, an intralayer exciton can absorb one hole from an interlayer trion, forming an intralayer trion and an interlayer exciton. The process lowers the total energy by $\varepsilon_3 - \varepsilon_t = 19.5$ meV > 15.5 meV. Therefore, the optically excited penton is unstable and will quickly dissociate into two parts after the photoexcitation, similar to the known dissociation process of short-lived negative molecular hydrogen ions $H_2^- \rightarrow H + H^-$ (*30*).

Discussion and outlook

Our experimental realization of a trion fluid in thermal equilibrium may lead to the observation of other exotic quantum phases. The charged nature of trions leads to strong intertrion Coulomb repulsion, and their heavy mass suppresses the kinetic energy. The dominance of Coulomb energy over kinetic energy can drive the system to spontaneously break the continuous translation symmetry and form a quantum crystal of trions (*15*, *33*, *34*). Recent theories have also predicted the possibility of interlayer trion–mediated topological superconductivity (*35*). In addition, we demonstrated Bose-Fermi mixtures composed of a tunable mixture of interlayer trions, interlayer excitons, and unpaired charges. As a condensed-matter analog of 2D molecules, these mixtures provide opportunities to explore quantum physics and chemistry with electrical tunability in the strongly correlated regime.

REFERENCES AND NOTES

1. M. A. Lampert, *Phys. Rev. Lett.* **1**, 450–453 (1958).
2. H. Wang *et al.*, *Phys. Rev. B* **93**, 045407 (2016).
3. P. Rivera *et al.*, *Nat. Commun.* **6**, 6242 (2015).
4. K. Kheng *et al.*, *Phys. Rev. Lett.* **71**, 1752–1755 (1993).
5. Z. Wang *et al.*, *Nature* **574**, 76–80 (2019).
6. L. Ma *et al.*, *Nature* **598**, 585–589 (2021).
7. R. Qi *et al.*, *Nat. Commun.* **14**, 8264 (2023).
8. M. M. Fogler, L. V. Butov, K. S. Novoselov, *Nat. Commun.* **5**, 4555 (2014).
9. F.-C. Wu, F. Xue, A. H. MacDonald, *Phys. Rev. B* **92**, 165121 (2015).
10. Y. Zeng, A. H. MacDonald, *Phys. Rev. B* **102**, 085154 (2020).
11. G. Wang *et al.*, *Rev. Mod. Phys.* **90**, 021001 (2018).
12. R. Qi *et al.*, *Science* **388**, 278–283 (2025).
13. P. X. Nguyen *et al.*, *Science* **388**, 274–278 (2025).
14. B. Debnath, Y. Barlas, D. Wickramaratne, M. R. Neupane, R. K. Lake, *Phys. Rev. B* **96**, 174504 (2017).
15. D. D. Dai, L. Fu, *Phys. Rev. Lett.* **132**, 196202 (2024).
16. A. Wójs, P. Hawrylak, J. J. Quinn, *Phys. Rev. B* **60**, 11661–11665 (1999).
17. A. Wójs, J. J. Quinn, P. Hawrylak, *Phys. Rev. B* **62**, 4630–4637 (2000).
18. R. Maezono, P. L. Ríos, T. Ogawa, R. J. Needs, *Phys. Rev. Lett.* **110**, 216407 (2013).
19. F. Cadiz *et al.*, *Phys. Rev. X* **7**, 021026 (2017).
20. C. Robert *et al.*, *Phys. Rev. Mater.* **2**, 011001 (2018).
21. Y. Zeng, V. Crépel, A. J. Millis, *Phys. Rev. Lett.* **132**, 266001 (2024).
22. O. Witham, R. J. Hunt, N. D. Drummond, *Phys. Rev. B* **97**, 075424 (2018).
23. See the supplementary materials.
24. K. F. Mak *et al.*, *Nat. Mater.* **12**, 207–211 (2013).
25. J. S. Ross *et al.*, *Nat. Commun.* **4**, 1474 (2013).
26. R. Tempelaar, T. C. Berkelbach, *Nat. Commun.* **10**, 3419 (2019).
27. D. Xiao, G.-B. Liu, W. Feng, X. Xu, W. Yao, *Phys. Rev. Lett.* **108**, 196802 (2012).
28. K. F. Mak, K. He, J. Shan, T. F. Heinz, *Nat. Nanotechnol.* **7**, 494–498 (2012).
29. K. F. Mak, D. Xiao, J. Shan, *Nat. Photonics* **12**, 451–460 (2018).
30. O. Heber *et al.*, *Phys. Rev. A* **73**, 060501 (2006).
31. C. Robert *et al.*, *Phys. Rev. Lett.* **126**, 067403 (2021).
32. I. Amelio, N. D. Drummond, E. Demler, R. Schmidt, A. Imamoglu, *Phys. Rev. B* **107**, 155303 (2023).
33. T. Smoleński *et al.*, *Nature* **595**, 53–57 (2021).
34. E. Wigner, *Phys. Rev.* **46**, 1002–1011 (1934).
35. C. Zerba, C. Kuhlenkamp, A. Imamoğlu, M. Knap, *Phys. Rev. Lett.* **133**, 056902 (2024).
36. R. Qi *et al.*, Dataset for "Electrically controlled interlayer trion fluid in electron-hole bilayers," Dryad (2025). https://doi.org/10.5061/dryad.m905qfvcm

ACKNOWLEDGMENTS

We thank I. Amelio, C. Robert, and X. Marie for their insightful discussion. **Funding:** The magneto-optic spectroscopy and data analysis were supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, under contract no. DE-AC02-05-CH11231 (van der Waals heterostructures program, KCWF16). The e-h bilayer device fabrication was supported by the Air Force Office of Scientific Research (AFOSR) award FA9550-23-1-0246. S.T. acknowledges primary support from DOE-SC0020653, Applied Materials Inc., NSF CMMI 1825594, NSF DMR-1955889, NSF CMMI-1933214, NSF 1904716, NSF 1935994, NSF ECCS 2052527, DMR 2111812, and CMMI 2129412. K.W. and T.T. acknowledge support from the JSPS KAKENHI (grant nos. 21H05233 and 23H02052) and World Premier International Research Center Initiative (WPI), MEXT, Japan for the growth of hBN crystals. R.Q. acknowledges support from the Kavli ENSI graduate student fellowship. **Author contributions:** F.W. conceived the research. R.Q. and Q.L. fabricated the devices with help from S.C., J.X., and Z.C. and supervision from A.Z. and F.W. R.Q. performed the optical measurements assisted by Z.Z., J.X., and A.Y.J. R.Q. and F.W. analyzed the data with inputs from D.D.D. and L.F. Y.O. and S.T. grew $MoSe_2$, MoS_2, and WSe_2 crystals. K.W. and T.T. grew hBN crystals. All authors discussed the results and wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data from this study are available at a Dryad repository (*36*).

License information: Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adn4584
Materials and Methods; Figs. S1 to S9

Submitted 11 December 2023; accepted 19 June 2025

10.1126/science.adn4584

2D MATERIALS

An equilibrium trion liquid in atomic double layers

Phuong X. Nguyen[1,2]†, Raghav Chaturvedi[1]†, Liguo Ma[1], Patrick Knuppel[1], Kenji Watanabe[3], Takashi Taniguchi[3], Kin Fai Mak[1,2,4]*, Jie Shan[1,2,4]*

A trion is a bound state of two electrons and one hole or two holes and one electron. To date, trions have been observed only as optically excited states in doped semiconductors. We report the emergence of an equilibrium trion liquid in Coulomb-coupled molybdenum diselenide ($MoSe_2$) and tungsten diselenide (WSe_2) monolayers. By electrically tuning the hole density in WSe_2 to be two times the electron density in $MoSe_2$, we generated equilibrium interlayer trions with binding energy in the milli–electron volt energy scale at temperatures two orders of magnitude below the Fermi temperature. We observed a density-tuned phase transition to an electron-hole plasma, spin-singlet correlations for the constituent holes, and Zeeman field–induced dissociation of the trions. Our results pave the way for exploration of the correlated phases of composite particles in solids.

Quantum liquids of coexisting oppositely charged particles are an ideal platform to generate quasiparticles with distinct quantum statistics and interactions and to realize proposed many-body phases of matter. Examples include exciton condensates (*1*–*18*), exciton-mediated superconductors (*19*–*22*), Wigner crystals of excitons and trions (*23*, *24*), and supersolids (*25*, *26*). Although it is straightforward to generate quantum fluids of electrons (e) or holes (h) in a solid-state system, their mixtures are intrinsically short-lived because of e-h recombination (*27*, *28*). The recent development of strongly Coulomb-coupled double layers of transition metal dichalcogenide (TMD) semiconductors (*14*, *15*, *17*, *18*, *29*, *30*) has overcome this experimental challenge. Metal electrodes can separately contact the two layers and inject electrons into one layer and holes into the other layer with independently tunable densities (*6*, *7*, *13*, *31*). This advance has enabled the thermodynamic (*15*, *30*) and transport studies (*17*, *18*) of strongly correlated excitons (bound e-h pairs, which are bosons) in thermal equilibrium. It has also opened the door to realize a Bose-Fermi mixture and new quasiparticles by adding electrons or holes to the exciton fluid (*32*, *33*).

Here, we report the emergence of an equilibrium trion liquid in Coulomb-coupled molybdenum diselenide ($MoSe_2$)/tungsten diselenide (WSe_2) double layers when the hole density (p) in the WSe_2 layer is two times the electron density (n) in the $MoSe_2$ layer (Fig. 1, A and B). By combining transport, capacitance, and optical measurements, we explored the properties of equilibrium trions (binding energy in the milli–electron volt energy scale and spin-singlet correlation for two constituent holes). We have also demonstrated a density-tuned quantum phase transition from a trion liquid to an e-h plasma as well as dissociation of trions by means of a Zeeman field of several teslas. Our results pave the way for the exploration of quantum many-body phenomena, including exciton-mediated pairing (*19*–*22*), unconventional trion and Fermi polaron metals (*34*), and trion-to-polaron crossover (*35*, *36*).

[1]School of Applied and Engineering Physics, Cornell University, Ithaca, NY, USA. [2]Kavli Institute at Cornell for Nanoscale Science, Ithaca, NY, USA. [3]National Institute for Materials Science, Tsukuba, Japan. [4]Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, NY, USA. ***Corresponding author. Email: kinfai.mak@cornell.edu (K.F.M.); jie.shan@cornell.edu (J.S.)** †These authors contributed equally to this work.

Measurement setup

We used the dual-gated device structure shown in Fig. 1C. The details have been reported previously in (*17*). The $MoSe_2$ (Mo) and WSe_2 (W) monolayers are separated by a hexagonal boron nitride (hBN) dielectric layer with a thickness of 1.5 to 2 nm in the channel. An interlayer bias voltage (V_b) tunes the e-h pair density (n_{pair}), with electrons residing in the Mo layer and holes in the W layer (Fig. 1A). The relatively thick hBN spacer, together with the intentionally angle-misaligned Mo and W layers, has been shown to effectively suppress interlayer tunneling to achieve equilibrium excitons and trions (fig. S8). At the same time, the spacer is sufficiently thin for the electrons and holes to be in the strong coupling regime. To complete the device, the e-h double layer is embedded in nearly symmetric top and bottom gates made of hBN dielectrics and few-layer graphite electrodes. The symmetric combination of the gate voltages, denoted as V_g, controls the net charge density or e-h imbalance; the antisymmetric combination, denoted as Δ, tunes the out-of-plane electric field, which can heavily dope the semiconductors in the contact region separated by a hBN spacer with a thickness of 10 to 20 nm and turn on the contacts.

We performed standard four-terminal resistance measurements on the W layer because it has lower contact resistances. The Mo layer was kept in the open-circuit configuration, and no current was allowed to flow in it (Fig. 1D). When the Mo layer is charge neutral ($n = 0$), the four-terminal resistance R probes the electrical transport properties of monolayer WSe_2. When the Mo layer is electron-doped, R becomes sensitive to the interlayer Coulomb interactions. Specifically, R is expected to exhibit characteristics of an insulator with a charge gap at low temperatures when interlayer bound states, such as trions, form. At elevated temperatures, R is inversely proportional to the thermally ionized hole density in the W layer; the charge gap thus characterizes the trion binding energy. We also measured the penetration capacitance (*15*) to verify that the trion fluid is charge compressible. We performed the magnetic circular dichroism (MCD) measurements to examine the spin configuration of the trions. We expected a spin-singlet correlation for the constituent holes to give rise to suppressed magnetic susceptibility of the W layer. We also expected an external Zeeman field to dissociate the trions by overcoming the spin gap and induce an insulator-to-metal transition in the W layer. We have studied three devices (figs. S1 and S3), and all results were reproduced. More details on the device fabrication and characterizations are shown in (*37*).

Trion liquid in thermal equilibrium

In Fig. 1, E and F, we show the penetration capacitance (C_P/C_{gg}; normalized by the gate-to-gate geometrical capacitance C_{gg}) and the four-terminal resistance R of the W layer in device 1 as a function of V_b and V_g at 1.5 K (Δ is fixed at 5.5 V). The penetration capacitance is proportional to the electronic incompressibility of the double layer (*15*). Specifically, $C_P/C_{gg} \approx 1$ corresponds to an incompressible state, and $C_P/C_{gg} \approx 0$ corresponds to a compressible state. Similar to earlier reports (*15*, *17*, *18*, *30*), we can identify four distinct regions in the electrostatic phase diagram for the W and Mo layers: *pn*, *pi*, *ii*, and *in*, where p, n, and i denote the hole-doped, electron-doped, and charge-neutral layers, respectively. The boundary between the *pi* and *ii* regions does not disperse with V_b because the W layer is grounded. We can also identify a charge-incompressible region of triangular shape in the *pn* region; it corresponds to the excitonic insulator phase with $n = p$.

The resistance R is largely consistent with the electrostatic phase diagram. We observed large resistance when the W layer was undoped (the *in*- and *ii*-regions) or had low hole densities. In the latter, holes were localized by disorders. We observed small resistance when the W layer was doped (the *pn*- and *pi*-regions), but with two exceptions. The first is the excitonic insulator phase. The observed large resistance is consistent with the recently reported perfect Coulomb drag and

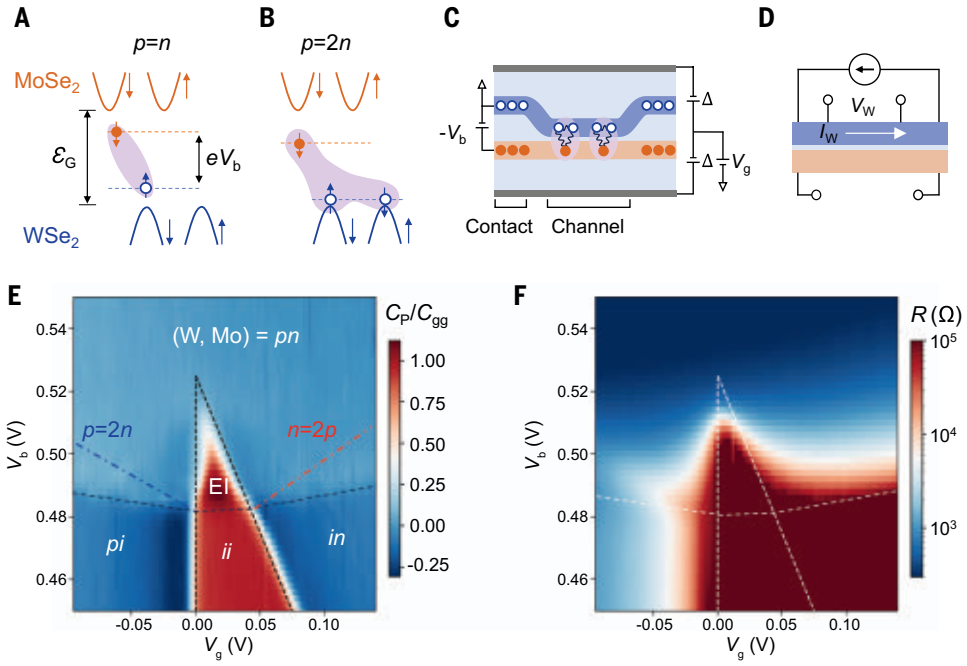


Fig. 1. Coulomb-coupled electron-hole double layers. (**A** and **B**) Schematic illustration of (A) excitons and (B) trions for one of the species. The conduction band minimum resides in the $MoSe_2$ layer, and the valence band maximum resides in the WSe_2 layer. Each band is spin-valley degenerate, with electron spins indicated with arrows. The positive trion is a spin-singlet state for two holes. Dashed lines are the chemical potential of electrons and holes. ε_G is the double-layer band gap, and eV_b is the exciton chemical potential. (**C**) Device schematic. The electron (orange) and hole (blue) layers are separated by 1.5 to 2 nm in the channel and 10 to 20 nm in the contact region. V_b, V_g, and Δ are the interlayer bias voltage and the symmetric and antisymmetric parts of the gate voltages, respectively. (**D**) Four-terminal resistance $R = V_W/I_W$ is measured in the hole layer with the electron layer in open circuit, where I_W and V_W are the bias current and voltage drop, respectively. (**E** and **F**) Normalized (E) penetration capacitance C_P/C_{gg} and (F) four-terminal resistance R versus V_g and V_b at 1.5 K (device 1). C_{gg} is the gate-to-gate geometrical capacitance. Dashed lines indicate the phase boundaries of hole-doped (p), electron-doped (n), and intrinsic (i) W and Mo layers. The electron-hole density ratio is 1:2 and 2:1 along the dash-dotted lines. EI, excitonic insulator.

diverging drag resistance from the formation of interlayer bound e-h pairs (*17*, *18*). The second exception is the state that stems from the lower left corner of the excitonic insulator phase and follows the $p = 2n$ line [density calibration is provided in fig. S2 and (*37*)]. This indicates the emergence of a distinct interlayer bound state.

We examined the temperature dependence of the state around $p = 2n$ (device 2) (Fig. 2). The state becomes more discernible in the resistance map at lattice temperature 0.02 K (Fig. 2A). We show R as a function of V_b at fixed V_g (Fig. 2A, red dashed line) for temperatures ranging from 0.02 to 10 K in Fig. 2B. At small V_b, we observed a V_b-independent resistance that exhibits a metallic behavior; that is, R increases with increasing temperature. When V_b exceeded a threshold of about 0.86 V, we observed a resistance drop followed by a sharp peak (Fig. 2B, shaded region). In this region, R displays an insulating behavior; that is, R increases with decreasing temperature. Upon further increases of V_b, R rapidly decreases and again exhibits a metallic behavior. The behavior outside the shaded region is understood from electrostatics. For the chosen V_g, the net charge density is fixed at $p - n \approx 2.42 \times 10^{11}$ cm^{-2}. Only when V_b exceeds the bias threshold can electron-hole pairs be injected into the bilayer. Therefore, for small V_b, the electron density in the Mo layer remained zero. The observed metallic behavior was expected for hole-doped monolayer WSe_2 with a constant hole density beyond the mobility edge ($p \approx 1.8 \times 10^{11}$ cm^{-2}) (fig. S7). Upon reaching the threshold, R drops owing to the increased hole density in the W layer. Again, for large V_b, the metallic behavior is accounted for by quick increases of hole density in the W layer.

The insulating behavior of the W layer upon an increase of the hole density (through injection of e-h pairs) in the shaded region of Fig. 2B cannot be explained by either single-particle physics or disorder. Disorder effects are inconsistent with the resistance drop

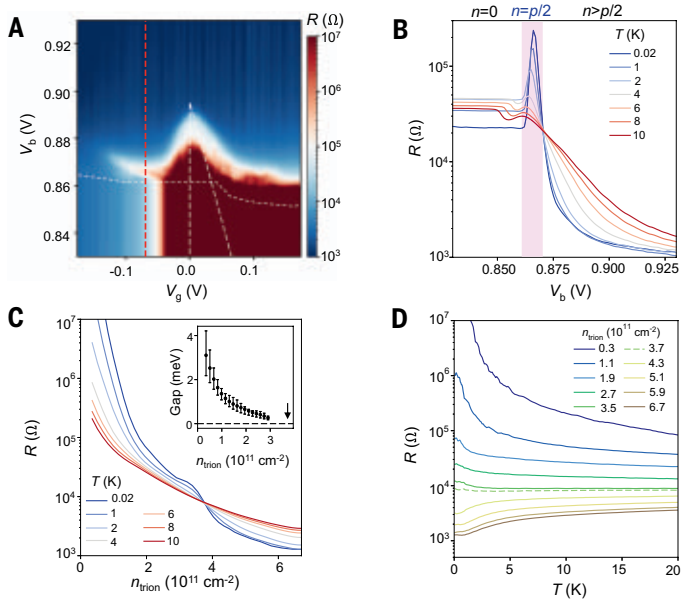


Fig. 2. Degenerate trion liquid. (**A**) Four-terminal resistance R versus V_b and V_g at lattice temperature 20 mK (device 2). White dashed lines indicate the electrostatic phase boundaries. (**B**) Linecut of (A) along the red dashed line at varying temperatures (0.02 to 10 K). (**C** and **D**) Four-terminal resistance R along the $p = 2n$ line (C) as a function of trion density at varying temperatures and (D) as a function of temperature at varying trion densities. (Inset) The estimated transport gap as a function of trion density. The vertical arrow indicates the critical trion density.

observed at slightly lower hole densities given that disorder potentials are expected to be more effectively screened in the shaded region (*38*). The insulating state emerges only around $p = 2n$. This shows that all the holes in the W layer are bound to the electrons in the Mo layer in a 2:1 ratio at low temperatures, providing strong evidence for the formation of positive trions.

We characterized the properties of trions by tracing the four-terminal resistance along the $p = 2n$ line. We show the trion density dependence of R at varying temperatures in Fig. 2C and the temperature dependence of R at varying trion densities in Fig. 2D. We observed a metal-insulator transition at critical trion density 3.75×10^{11} cm^{-2}, at which R is nearly temperature independent (Fig. 2D, dashed line). The system is insulating below and metallic above the critical density. We estimated the charge gap of the insulating state by performing a thermal activation fit to the temperature dependence of R (fig. S4). The density dependence of the gap size is shown in Fig. 2C, inset. The charge gap decreases monotonically with increasing trion density and vanishes continuously at the critical density.

The charge gap is a measure of the trion binding energy because R is inversely proportional to the thermally ionized hole density. It is in the 0- to 3-meV energy range in comparison with 20 to 30 meV for intralayer trions in TMD monolayers (*28*). The extracted trion binding energy is up to a few percent of the exciton binding energy of ≈40 meV (figs. S8, S9, and S11), which is consistent with the expected ratio for trion-to-exciton binding (*39*–*41*). The corresponding trion Fermi temperature is estimated to be ≈5 K (*37*). The equilibrium trion liquid is thus at a temperature as low as two orders of magnitude below the Fermi temperature. The trion liquid is also charge compressible, as shown by the penetration capacitance result (Fig. 1E); any injected holes are readily accepted by the exciton reservoir to form trions.

Our result (Fig. 2C) further suggests a density-tuned continuous phase transition from a trion liquid to an e-h plasma at the critical trion density, at which the trion binding energy vanishes owing to free carrier screening. Because the trion critical density (≈3.75×10^{11} cm^{-2}) is not substantially higher than the hole mobility edge, future studies are required to address whether the trions are delocalized near the phase transition. In the plasma phase, excitons are unlikely to exist because the trion critical density is comparable with the exciton Mott density of ≈4×10^{11} cm^{-2} (*17*). This suggests that exciton dissociation is the limiting factor for trion formation, a point also worthy of further investigation.

We also expected the negative trion state to emerge from the lower right corner of the excitonic insulator phase around $n = 2p$. This was not observed in the current transport data because of the large resistance background. The W layer is highly resistive in this region because of the overall low hole densities and strong excitonic binding of the holes to the electrons in the Mo layer. The negative trion state could be best accessed with four-terminal resistance measurements on the Mo layer. We observed signatures of the negative trion state in the complementary optical measurement (fig. S6).

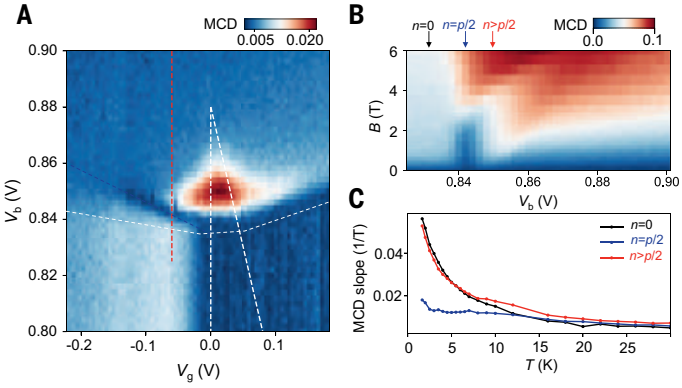


Fig. 3. Spin-singlet correlation in trions. (**A**) Spectrally integrated MCD of the W layer as a function of V_b and V_g (device 3; out-of-plane magnetic field $B = 0.2$ T, temperature 1.6 K). White dashed lines indicate the electrostatic phase boundaries. MCD is suppressed around the $p = 2n$ region (blue dashed line). (**B**) MCD versus V_b and B along the red dashed line in (A) (the corresponding net charge density is approximately $+1.9 \times 10^{11}$ cm^{-2}). (**C**) Temperature dependence of the zero-field MCD slope at $V_b = 0.83$, 0.842, and 0.85 V, corresponding to $n = 0$, $n = p/2$, and $n > p/2$, respectively. The quantity is proportional to the magnetic susceptibility of the W layer. Susceptibility at $p = 2n$ is suppressed below about 10 K.

Spin-singlet correlation in trions

Next, we examined the spin configuration of trions by performing magneto-optical measurements (fig. S6) (*37*). In Fig. 3A, we show the magnetic circular dichroism (MCD) of the W layer in device 3 as a function of V_b and V_g under out-of-plane magnetic field $B = 0.2$ T at 1.6 K. The MCD (spectrally integrated over the intralayer charged exciton resonance) has been shown to be proportional to the spin (locked with valley) polarization of doped carriers in monolayer TMDs (*42*). In the small field limit, MCD is proportional to the magnetic susceptibility (*42*). We observed notably enhanced magnetic susceptibility of the W layer in the excitonic insulator phase. We also observed a susceptibility enhancement for small hole densities (boundary between the *pi* and *ii* regions), where the hole band mass and *g*-factor were renormalized owing to strong interactions (*43*, *44*). The most intriguing observation is the susceptibility suppression around $p = 2n$, which fades away with increasing trion density above the critical value determined from the transport study.

In Fig. 3C, we display the temperature dependence of susceptibility for three representative values of V_b at a fixed V_g (Fig. 3A, red dashed line), corresponding to $n = 0$, $n = p/2$, and $n > p/2$, respectively. At high temperatures ($T \gtrsim 10$ K), the susceptibility is comparable in all three cases. At low temperatures, the susceptibility for the case of $p = 2n$ is strongly suppressed compared with the other two cases, for which the susceptibility rapidly increases with decreasing temperature. In Fig. 3B, we show MCD along the red dashed line in Fig. 3A as a function of B up to 6 T. The MCD around $p = 2n$ remains small up to about 3 to 4 T, above which it rapidly saturates.

The substantially suppressed low-temperature susceptibility around $p = 2n$ demonstrates a spin-singlet correlation for the two holes in a positive trion, which resists spin-valley alignment by a small external magnetic field. This is

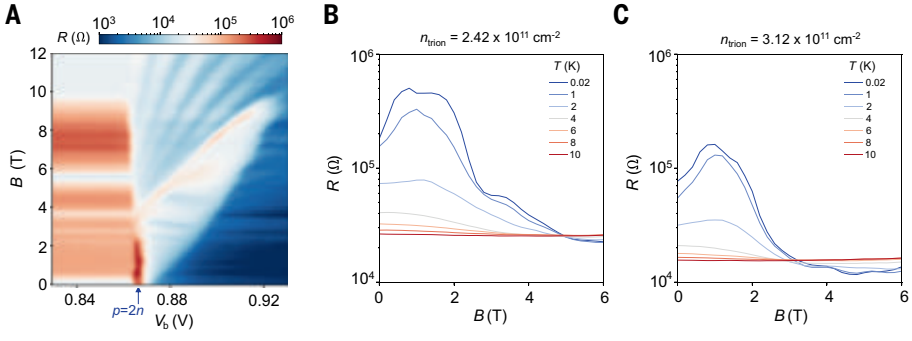


Fig. 4. Zeeman field–induced dissociation of trions. (**A**) Four-terminal resistance R versus V_b and B at constant V_g at 20 mK (device 2). The corresponding net-positive charge density is approximately 2.4×10^{11} cm^{-2}. (**B**) Vertical linecut of (A) along $p = 2n$ at varying temperatures. The temperature dependence of R is characteristic of an insulator below and of a metal above about 5 T. (**C**) Same as (B) at a fixed trion density of 3.1×10^{11} cm^{-2}. The critical field decreases with increasing trion density.

consistent with earlier optical studies of monolayer TMDs, in which the ground state trions are a spin-singlet (two holes occupy different valley-spin states) (*28*, *45*). The spin-singlet correlation is important only below the trion ionization temperature when a substantial population of trions is stabilized. This temperature scale (10 K) is consistent with the trion binding energy (1 meV) at $n_{\mathrm{trion}} \approx 1.9 \times 10^{11}\ \mathrm{cm}^{-2}$. The spin-singlet correlation can also be overcome when the Zeeman energy associated with B becomes comparable with the trion binding energy. The observed Zeeman field scale (3 to 4 T) is also consistent with the trion binding energy (Fig. 2C, inset), using the reported hole g-factor $g \approx 10$ (*43*).

Zeeman field–induced dissociation of trions

Last, we examined the fate of the trion liquid after the Zeeman field overcomes the spin-singlet correlation. In Fig. 4A, we show four-terminal resistance R of the W layer as a function of B and V_{b} at a fixed V_{g}, which corresponds to approximately constant net charge density $p - n \approx 2.4 \times 10^{11}\ \mathrm{m}^{-2}$. We observed three distinct sets of quantum oscillations. The horizontal lines in Fig. 4A below the pair injection threshold ($n = 0$ or $V_{\mathrm{b}} < 0.866$ V) are the quantum oscillations of monolayer WSe_2. They are observable above about 2 T, which corresponds to a quantum mobility exceeding 5000 $\mathrm{cm}^2\ \mathrm{V}^{-1}\ \mathrm{s}^{-1}$. The nearly horizontal lines in Fig. 4A at large V_{b} may arise from trions and require further studies. The main feature of interest is the Landau fan that emerges from $V_{\mathrm{b}} = 0.866$ V (corresponding to $p = 2n$) above a critical field of about 5 T. The Landau fan disperses with both V_{b} and B.

In Fig. 4B, we display the magnetic-field dependence of R for $p = 2n$ at varying temperatures. We observed a magnetic field–induced metal-insulator transition at the critical field. The system is insulating below and metallic above the critical field. The critical field agrees well with the field for magnetic saturation in Fig. 3B. The same study for a higher trion density is shown in Fig. 4C. We observed the same properties but at a lower critical magnetic field.

The metal-insulator transition shows that once the Zeeman field overcomes the spin-singlet correlation and polarizes all holes to a single valley (spin) in the W layer, the trion liquid dissociates into an e-h plasma. This gives rise to a Landau fan for the ionized holes that disperses with both V_{b} and B for $V_{\mathrm{b}} \gtrsim 0.866$ V and $B \gtrsim 5$ T. The Zeeman field also reduces the effective interlayer band gap of the bilayer, resulting in a small shift of the $p = 2n$ line to smaller V_{b} in Fig. 4A. The picture of Zeeman field–induced trion dissociation is further supported by the observed smaller critical field for the higher trion density (Fig. 4, B and C), given that the trion binding energy decreases with increasing density (Fig. 2C, inset) and a smaller trion binding energy requires a smaller Zeeman field to overcome the spin-singlet correlation. The result is inconsistent with partial ionization for a mixture of holes and excitons because in this scenario, the hole Landau levels are not expected to disperse with V_{b} (which tunes the e-h pair density). Complete ionization of trions is presumably caused by the substantial state-filling and free-carrier screening effects (*28*, *45*) after magnetic saturation. Further theoretical studies are required to fully understand the field-induced trion dissociation process.

Conclusions and outlook

We observed the emergence of an equilibrium trion liquid in Coulomb-coupled $MoSe_2/WSe_2$ e-h double layers at $p = 2n$. The trion liquid is compressible with a trion binding energy in the milli–electron volt energy scale. Magneto-optical studies showed that the two holes in a positive trion have spin-singlet correlation. We further demonstrated density-tuned and magnetic field–tuned quantum phase transitions from a trion liquid to an e-h plasma. Our results open the door for exploring distinct transport phenomena of trions, such as the trion quantum oscillations and trion Hall effect, as well as quantum many-body states of trions, such as trion Wigner crystals and trion Cooper pairs.

REFERENCES AND NOTES

1. Y. E. Lozovik, V. Yudson, *Zh. Eksp. Teor. Fiz* **71**, 738 (1976).
2. X. Zhu, P. B. Littlewood, M. S. Hybertsen, T. M. Rice, *Phys. Rev. Lett.* **74**, 1633–1636 (1995).
3. J. P. Eisenstein, A. H. Macdonald, *Nature* **432**, 691–694 (2004).
4. L. V. Butov, *J. Phys. Condens. Matter* **16**, R1577–R1613 (2004).
5. J.-J. Su, A. H. MacDonald, *Nat. Phys.* **4**, 799–802 (2008).
6. M. M. Fogler, L. V. Butov, K. S. Novoselov, *Nat. Commun.* **5**, 4555 (2014).
7. F.-C. Wu, F. Xue, A. H. MacDonald, *Phys. Rev. B Condens. Matter Mater. Phys.* **92**, 165121 (2015).
8. L. Du *et al.*, *Nat. Commun.* **8**, 1971 (2017).
9. J. I. A. Li, T. Taniguchi, K. Watanabe, J. Hone, C. R. Dean, *Nat. Phys.* **13**, 751–755 (2017).
10. X. Liu, K. Watanabe, T. Taniguchi, B. I. Halperin, P. Kim, *Nat. Phys.* **13**, 746–750 (2017).
11. G. W. Burg *et al.*, *Phys. Rev. Lett.* **120**, 177702 (2018).
12. P. López Ríos, A. Perali, R. J. Needs, D. Neilson, *Phys. Rev. Lett.* **120**, 177701 (2018).
13. M. Xie, A. H. MacDonald, *Phys. Rev. Lett.* **121**, 067702 (2018).
14. Z. Wang *et al.*, *Nature* **574**, 76–80 (2019).
15. L. Ma *et al.*, *Nature* **598**, 585–589 (2021).
16. X. Liu *et al.*, *Science* **375**, 205–209 (2022).
17. P. X. Nguyen *et al.*, *Science* **388**, 274–278 (2025).
18. R. Qi *et al.*, *Science* **388**, 278–283 (2025).
19. O. Cotleţ, S. Zeytinoğlu, M. Sigrist, E. Demler, A. Imamoğlu, *Phys. Rev. B* **93**, 054510 (2016).
20. V. Crépel, L. Fu, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2117735119 (2022).
21. C. Zerba, C. Kuhlenkamp, A. Imamoğlu, M. Knap, *Phys. Rev. Lett.* **133**, 056902 (2024).
22. V. Crépel, D. Guerci, J. Cano, J. H. Pixley, A. Millis, *Phys. Rev. Lett.* **131**, 056001 (2023).
23. M. A. Baranov, M. Dalmonte, G. Pupillo, P. Zoller, *Chem. Rev.* **112**, 5012–5061 (2012).
24. D. D. Dai, L. Fu, *Phys. Rev. Lett.* **132**, 196202 (2024).
25. Y. N. Joglekar, A. V. Balatsky, S. Das Sarma, *Phys. Rev. B Condens. Matter Mater. Phys.* **74**, 233302 (2006).
26. S. Conti *et al.*, *Phys. Rev. Lett.* **130**, 057001 (2023).
27. K. F. Mak, J. Shan, *Nat. Photonics* **10**, 216–226 (2016).
28. G. Wang *et al.*, *Rev. Mod. Phys.* **90**, 021001 (2018).
29. Y. Zhou *et al.*, *Nature* **595**, 48–52 (2021).
30. R. Qi *et al.*, *Nat. Commun.* **14**, 8264 (2023).
31. Y. Zeng, A. H. MacDonald, *Phys. Rev. B* **102**, 085154 (2020).
32. A. Wójs, P. Hawrylak, J. J. Quinn, *Phys. Rev. B Condens. Matter* **60**, 11661–11665 (1999).
33. A. Wójs, J. Quinn, P. Hawrylak, *Phys. Rev. B Condens. Matter* **62**, 4630–4637 (2000).
34. M. Sidler *et al.*, *Nat. Phys.* **13**, 255–261 (2017).
35. M. Duda *et al.*, *Nat. Phys.* **19**, 720–725 (2023).
36. L. De Marco *et al.*, *Science* **363**, 853–856 (2019).
37. Materials and methods are available as supplementary materials.
38. B. Spivak, S. V. Kravchenko, S. A. Kivelson, X. P. A. Gao, *Rev. Mod. Phys.* **82**, 1743–1766 (2010).
39. E. V. Calman *et al.*, *Nano Lett.* **20**, 1869–1875 (2020).
40. T. Deilmann, K. S. Thygesen, *Nano Lett.* **18**, 1460–1465 (2018).
41. O. Witham, R. J. Hunt, N. D. Drummond, *Phys. Rev. B* **97**, 075424 (2018).
42. K. F. Mak, J. Shan, *Nat. Nanotechnol.* **17**, 686–695 (2022).
43. Z. Wang, J. Shan, K. F. Mak, *Nat. Nanotechnol.* **12**, 144–149 (2017).
44. Q. Shi *et al.*, *Nat. Nanotechnol.* **15**, 569–573 (2020).
45. T. Smoleński *et al.*, *Phys. Rev. Lett.* **123**, 097403 (2019).
46. P. X. Nguyen *et al*., An equilibrium trion liquid in atomic double layers. Zenodo (2025); https://doi.org/10.5281/zenodo.15670774.

ACKNOWLEDGMENTS

We thank A. H. MacDonald, Y. Zeng, and B. Zou for helpful discussions. **Funding:** This work was supported by the US National Science Foundation (NSF) under DMR-2004451 (capacitance and optical measurements) and the US Department of Energy (DOE), Office of Science, Basic Energy Sciences (BES), under award DE-SC0022058 (electrical measurement). It was also funded in part by the Gordon and Betty Moore Foundation (grant GBMF11563). Growth of the hBN crystals was supported by the Elemental Strategy Initiative of MEXT, Japan, and CREST (JPMJCR15F3), JST. This work was performed in part at the Cornell NanoScale Facility, an NNCI member supported by NSF grant NNCI-2025233. We also acknowledge support from the David and Lucile Packard Fellowship (to K.F.M.) and Kavli Institute at Cornell (KIC) Engineering Graduate Fellowship (to P.X.N.). **Author contributions:** P.X.N. and R.C. fabricated the devices, performed the measurements, and analyzed the data. L.M. contributed to the device fabrication and data analysis. P.K. contributed to the setup and data analysis of the optical spectroscopy measurements. K.W. and T.T. grew the bulk hBN crystals. K.F.M. and J.S. designed the scientific objectives and oversaw the project. All authors discussed the results and commented on the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data shown in the main text and supplementary materials are available in the Zenodo data repository (*46*).

License information: Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adn5886
Materials and Methods; Supplementary Text; Figs. S1 to S11; References (*47*–*53*)

Submitted 19 December 2023; accepted 29 August 2025

10.1126/science.adn5886

ScienceCareers.org
ScienceCareers

UHealth UNIVERSITY OF MIAMI HEALTH SYSTEM | UNIVERSITY OF MIAMI MILLER SCHOOL of MEDICINE

Multiple Tenure-Track Faculty Positions
(Assistant, Associate, and Full Professor)

Department of Cell and Systems Biology

The Department of Cell and Systems Biology at the Miller School of Medicine | University of Miami is expanding and recruiting bold scientists who want to reshape the future of biomedicine. We invite applications for tenure-track/tenured faculty positions at all ranks. We seek exceptionally innovative investigators (fundamental or applied) whose research is visionary, disruptive, and pushes boundaries in fields such as quantitative systems and synthetic biology, development, aging and cancer, infectious disease, and the evolution of resistance. **Our ethos prioritizes applicants with highly creative, non-derivative research programs over level of experience and research background**. Investigators with translational or entrepreneurial interests are encouraged to apply.
A Ph.D., M.D. or equivalent, is required and both research and teaching are expected.

The Miller school offers state-of-the-art research facilities, robust start-up packages tailored to ensure your success, and premier salary support, commensurate with experience. Please send a CV, statement of research accomplishments, future research goals, three representative publications or preprints, and the names of three references to: rbraithw@med.miami.edu.

We are building a new generation of discovery in Miami — join us.
For further information: https://sysbio.miami.edu/

UHealth UNIVERSTIY OF MIAMI HEALTH SYSTEM

MAX PLANCK INSTITUTE
FOR MULTIDISCIPLINARY SCIENCES

The Max Planck Institute for Multidisciplinary Sciences is a leading international research institute of exceptional scientific breadth. With more than 40 research groups and some 1,000 employees from over 50 nations, it is the largest institute of the Max Planck Society.

The department *Dynamics at Surfaces* (Prof. Dr. Alec M. Wodtke) is inviting applications for a position as

Postdoctoral researcher (f/m/d)

We anticipate a start date between January and March of 2026. The anticipated duration of the project is 3 years. Payment and benefits are based on the German Public Service Payscale (TVöD Bund) guidelines.

The Max Planck Society is committed to increasing the number of individuals with disabilities in its workforce and, therefore, encourages applications from such qualified individuals. The Max Planck Society strives for diversity and gender equality and welcomes applications from all backgrounds.

Please send your application including a cover letter (explaining background and motivation), CV, transcript of records, contact names and addresses of two to four referees familiar with the work of the candidate preferably via e-mail as one single PDF file **until 02.11.2025** to:
ausschreibung27-25@mpinat.mpg.de

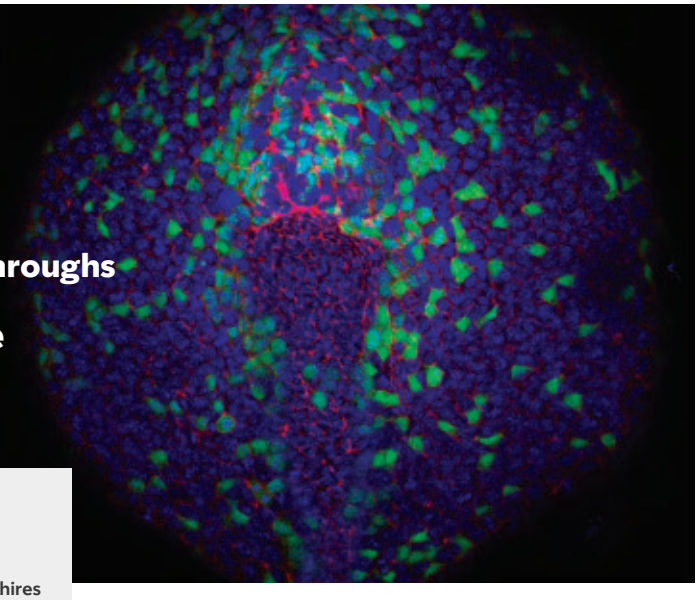
Max Planck Institute for Multidisciplinary Sciences
Department of Dynamics at Surfaces
Prof. Dr. Alec M. Wodtke
Prof. Dr. Dirk Schwarzer
Am Fassberg 11
37077 Göttingen
Germany

More information about the job advertisement can be found here: www.mpinat.mpg.de/27-25



Sloan Kettering Institute

Celebrating over 75 years of breakthroughs



Join our faculty and help set the pace of biomedical science.

Join Us

Successful candidates will hold an appointment in one of SKI's research programs. Candidates may apply to up to two programs. Salary range is $195K-260K. Starting pay for new hires typically falls within the range shown above, based on candidate qualifications and experience, as well as other variables.This range represents annual salary only and does not include any one-time payments that eligible candidates may be offered at the time of hire.

Cancer Biology & Genetics
Chair: Scott Lowe, PhD

Cell Biology
Chair: Tobias Walther, PhD

Chemical Biology
Chair: Derek Tan, PhD

Computational & Systems Biology
Chair: Dana Pe'er, PhD

Developmental Biology
Chair: Anna-Katerina Hadjantonakis, PhD

Human Oncology and Pathogenesis Program
Chair: Charles Sawyers, MD

Immunology
Chair: Alexander Rudensky, PhD

Immuno-Oncology
Chair: Andy Minn, MD, PhD

Molecular Biology
Chair: John Petrini, PhD

Molecular Pharmacology
Chair: Omar Abdel-Wahab, MD

Structural Biology
Chair: Christopher Lima, PhD



Memorial Sloan Kettering Cancer Center

Areas of Basic and Translational Research

- Stem Cell Biology
- Developmental Oncology
- Regenerative Medicine
- Machine Learning
- 3D Single-Cell Analytics
- Biophysics & Imaging
- Organelle Biology
- Chromatin & Gene Regulation
- Genome Integrity and Functional Genomics
- Chemistry & Chemical Biology
- Metastasis & Tumor Microenvironment
- Immunity, Host-Microbial Interactions and Microbiomes
- Tumor Immunobiology
- Experimental Therapeutics, Imaging and Bioengineering
- Immuno-oncology

Research and Training

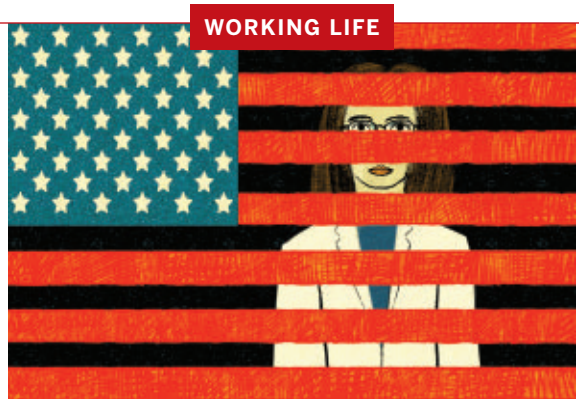
- 120 research laboratories housed in state-of-the-art buildings
- 26 Core facilities offering cutting-edge scientific services
- More than 900 pre- and postdoctoral trainees
- Appointments in the Gerstner Sloan Kettering Graduate School of Biomedical Sciences and the Weill Cornell Graduate School of Medical Sciences

Visit **www.ski.edu** to learn more.

MSK is an equal opportunity and affirmative action employer committed to diversity and inclusion in all aspects of recruiting and employment. All qualified individuals are encouraged to apply and will receive consideration without regard to race, color, gender, gender identity or expression, sexual orientation, national origin, age, religion, creed, disability, veteran status or any other factor which cannot lawfully be used as a basis for an employment decision.

Application deadline: **November 15, 2025** | Apply at: **https://facultysearch.ski.edu**

ScienceCareers.org
Science Careers



An immigrant's dilemma

Mariana Gelambi

On a warm June evening, I sat alone in my house, sobbing as I watched a choppy, pixelated livestream of my mother's funeral. Thousands of kilometers away in Venezuela, my family grieved together while I remained in the United States, unable to travel because of immigration restrictions. I tried to tell my 17-month-old son I was sad, but not because of him. I was heartbroken and overwhelmed with guilt for not being there. I kept asking myself whether choosing a career in science, so far from home, had been the right decision. I felt a little like the species I study as a postdoc: the spotted lanternfly, an invasive insect that threatens U.S. ecosystems. I, too, was unwelcome and out of place.

As an undergraduate in Venezuela, I became captivated by the biodiversity of the tropics and began to research fruit bats. A cold email to a U.S. scientist working in the field opened the door for me to pursue a Ph.D. in her lab. I worried about not fitting in and was sad to leave home. Still, I felt thrilled and fortunate to move abroad. Visiting my family back home would never be a problem, or so I assumed. And after completing my Ph.D., I could return to Venezuela, teach at a local university, and share what I had learned abroad.

During my first year as a Ph.D. student, my mother was diagnosed with breast cancer. My first instinct was to move back home, but my mother—wise and selfless—assured me I should stay and keep pursuing my dream. Her illness made me more determined to finish so the sacrifice would feel worthwhile. I visited her three times during my program, and we shared long video calls. I followed her doctors' visits, her hair loss, her fading strength. She supported me amid my research struggles and celebrated my progress, professionally and personally. She met my partner virtually. She watched my Ph.D. graduation online, and my wedding 2 months later. Despite my deep sadness, I found joy in my work and felt deeply grateful for the life I was building in the U.S.

Meanwhile my plan to return home had begun to dissolve, as Venezuela's political and economic collapse made academic careers nearly impossible there. After my marriage I submitted a green card application so I could continue my career in the U.S. In the meantime, I transitioned to a postdoctoral position studying invasive species, working under an Optional Practical Training extension—a program that allows student visa holders in STEM to work in the U.S. for up to 2 years after finishing their degree.

Still, I wanted to visit Venezuela. My husband and I had welcomed our first son, and my biggest dream was to take him to meet my mother; her cancer had metastasized to her bones, leaving her too weak to travel. But until I had my green card, I could not leave the U.S. Under my student visa extension, I would not be allowed to re-enter the country. My husband and I spent countless hours navigating the immigration system, reading confusing guidelines, filling out dozens of forms, triple checking instructions. I lived in a constant state of waiting, hoping my green card would be approved.

In June, new regulations and a travel ban triggered warnings from my university's international office, warning that my visa could be revoked at any moment and further discouraging international travel. I was advised to speak with immigration attorneys. For the first time, I began to question whether I truly belonged in the U.S. I even became self-conscious about my accent and appearance.

The clock was ticking. My mother's health was deteriorating rapidly. Each morning, I checked my green card status. I was desperate to see her one more time. But the most consistent advice remained: Do not leave. A few weeks later, my mother passed away. I never got to say goodbye, and she never met my son.

I find comfort in believing my mother was proud that the curious girl she raised had followed her dream of becoming a scientist and turned that curiosity into a career. And although some may see me, and others like me, as invasive pests, I choose to focus on a different aspect of my study species: Even in foreign lands, it's possible to survive, adapt, and thrive. □

Mariana Gelambi is a postdoc at the Virginia Polytechnic Institute and State University. Send your career story to SciCareerEditor@aaas.org.

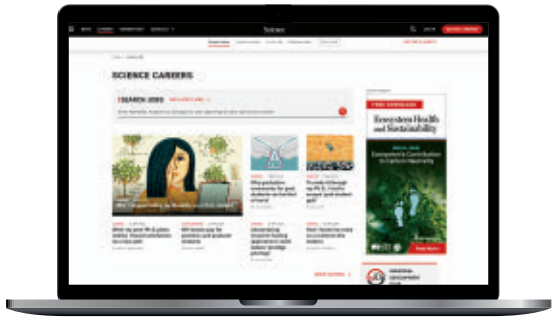
ILLUSTRATION: ROBERT NEUBECKER



Let *Science* Careers help advance your career.

- Register for a free online account on **ScienceCareers.org**.
- Search hundreds of job postings.
- Sign up to receive job alerts that match your criteria.
- Upload your resume into our database to connect with employers.
- Watch one of our many webinars on different career topics such as job searching, networking, and more.
- Download our career booklets, including Career Basics, Careers Beyond the Bench, and Developing Your Skills.
- Complete a personalized career plan at "my IDP."
- Read relevant career advice articles from our library of thousands.

Visit **ScienceCareers.org** today — all resources are free



Science Careers
FROM THE JOURNAL SCIENCE AAAS

SCIENCECAREERS.ORG



BII | Prize for Science | Innovation

Bring Science to Life

Behind every life-changing solution is an entrepreneurial scientist – a creative mind who developed an idea in the lab and dared to carry it out into the world.

To encourage more scientists to translate their research, BioInnovation Institute (BII) and *Science* are collaborating to host an annual award.

Up to three winners will have their essays published in *Science* magazine and will be invited into BII's entrepreneurial ecosystem. In addition, the Grand Prize winner will receive a prize of USD 25,000, and each finalist will receive USD 10,000 at a grand award ceremony in Copenhagen, Denmark.

Xiaojing Gao, Grand Prize winner 2025
For his work on developing an innovative programmable RNA sensor and molecular tools to improve mRNA vector therapeutics.

BII BioInnovation Institute Science

Apply before November 1, 2025
bii.dk/community/prizes