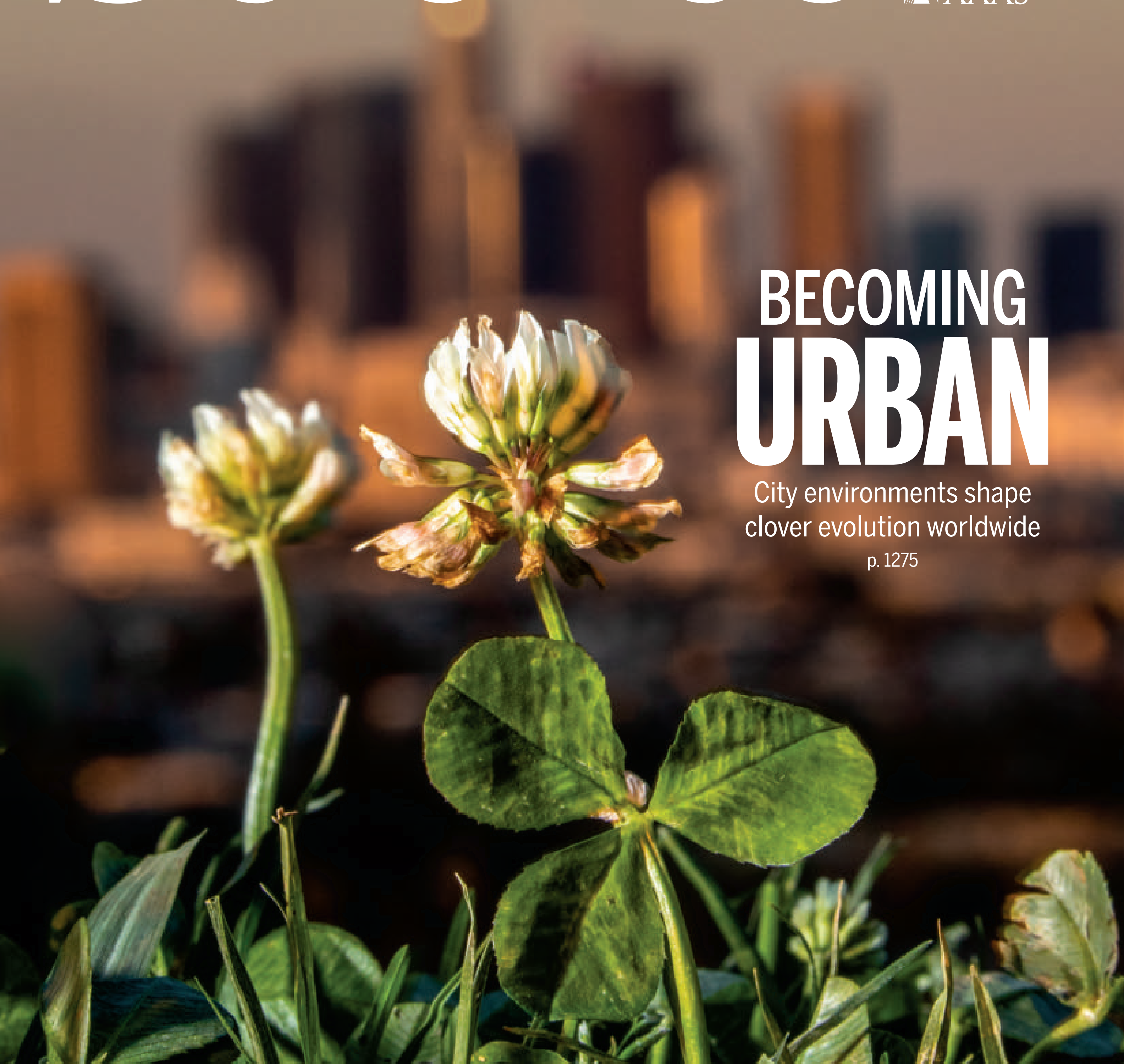
Scientists seek refuge from war in Ukraine p. 1209

Burning up nutrients gets fat cells hot p. 1229

Large wildfires change stratospheric chemistry p. 1292

Science

$15
18 MARCH 2022
science.org
AAAS



BECOMING URBAN

City environments shape clover evolution worldwide

p. 1275

The Harmonious Coexistence between Nature and Humankind

The International Cosmos Prize

To carry on and further develop the philosophy of "The Harmonious Coexistence between Nature and Humankind" presented at the International Garden and Greenery Exposition, Osaka, Japan, 1990 (Expo '90), the International Cosmos Prize recognizes outstanding research activities and achievements both in Japan and abroad that are in line with this philosophy.
The award ceremony is held every autumn in Osaka, Japan. The prizewinner is awarded a monetary prize of 40 million yen.

2021 Prizewinner

Dr. Peter Bellwood

Emeritus Professor, Australian National University

Dr. Peter Bellwood is a prominent archaeologist, who has been exploring the process of agricultural dispersal—which has been closely correlated with human migrations—from a global viewpoint, while studying human life in Oceania and Southeast Asia during the Neolithic Age as his main research theme. He has proposed the "early farming dispersal hypothesis" based on interdisciplinary research in archaeology, linguistics and human biology, clarifying the agricultural origins and the process of early farmer migration and dispersal.

Prizewinners 2015-2019

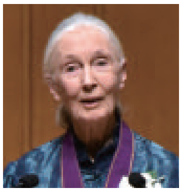
2015

Dr. Johan Rockström

2016

Dr. IWATSUKI Kunio

2017



Dr. Jane Goodall

2018

Dr. Augustin Berque

2019

Dr. Stuart L. Pimm

- 2020 award were cancelleddue to COVID-19 pandemic.

For inquiries about the International Cosmos Prize, please send an email from this E-mail adress.
EXPO '90 FOUNDATION E-mail: rec-cosmos@expo-cosmos.or.jp

Advertorial

Skin in the game: Procter & Gamble scientists identify and target a novel pathway involved in hyperpigmented skin spots in vitro.

Procter & Gamble scientists discover novel pathway causing skin hyperpigmentation

Every day, people around the world look in a mirror and notice a new speckle, dot, or splotch on their skin. Spots caused by hyperpigmentation—where an area of skin is darkened—affect all ethnicities, skin tones, genders, and ages, and can be a point of pride, frustration, or perplexity for the owner. For decades, scientists at The Procter & Gamble (P&G) Company, which owns several of the world's leading skincare brands including OLAY, have been trying to unravel the mystery of how hyperpigmented spots form and grow. Now they are closer to cracking a new code, thanks in part to the creativity of one scientist and a supporting cast of thousands of other scientists and engineers.

Tomohiro Hakozaki, a research fellow in P&G's Beauty Technology Division, has been singularly focused on skin pigmentation for almost his entire career. With a Ph.D. in dermatological science, Hakozaki has worked in various areas of skin biology and biometrics as he has sought to better comprehend the mechanism within skin cells that causes hyperpigmentation. "You might remember when you first see the spot on your face, and you react 'what's this?'" he says. "When we do consumer research, spots always pop up as a number one or two skincare concern. This is what motivates my life's work. My dream is to develop the technology to fade and ultimately remove unwanted spots."

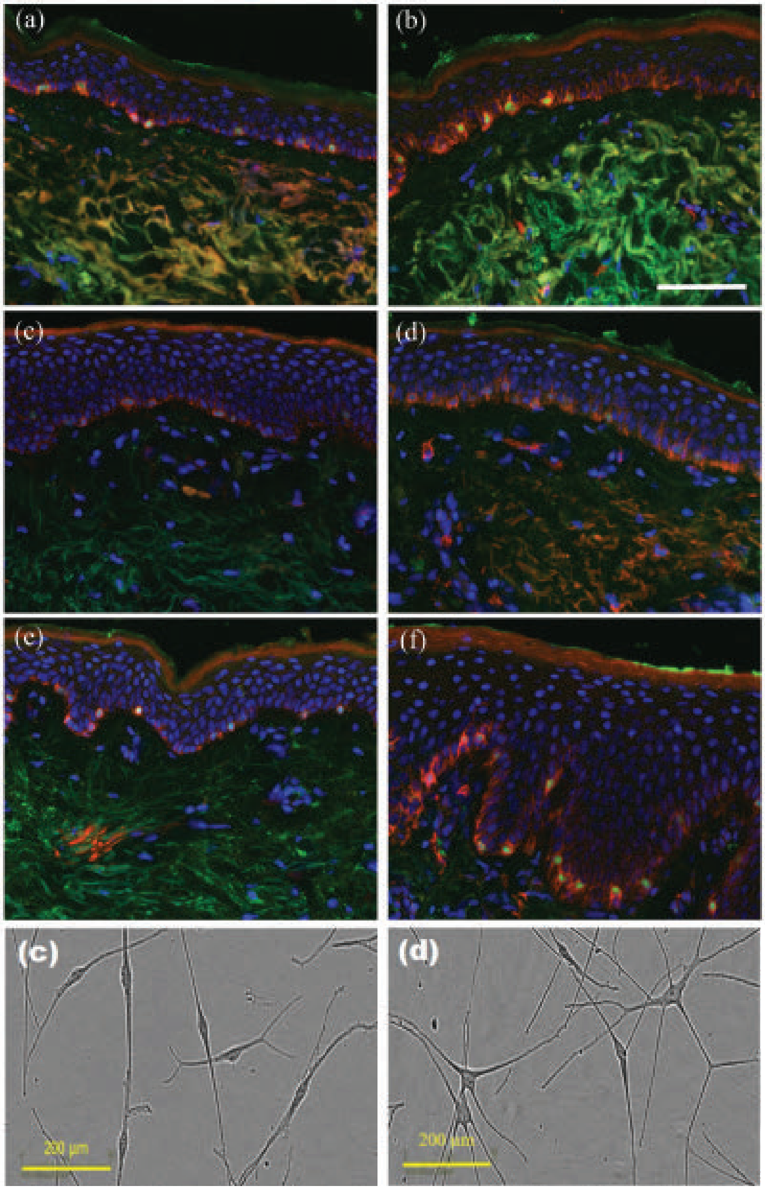
To find a solution to stop the spots, P&G scientists leveraged their knowledge about skin and pigmentation. While human skin color has several components, the biggest driver is melanin, and when it is localized in some areas, it appears as a spot. Melanin is produced in melanocytes, very specialized skin cells with dendritic properties: In the human epidermis, one dendritic melanocyte interacts with about 36 keratinocytes (another type of skin cell) and supplies them with melanin. We know that the presence of hyperpigmented spots is influenced by multiple intrinsic and extrinsic factors, such as ultraviolet B (UVB) exposure, hormone imbalance, inflammatory status, and aging; there has also been abundant research on the factors influencing melanin pathways, which include melanin synthesis, melanogenic cytokines, and melanosome transfer. However, the precise mechanism underlying melanocyte dendricity, or branching, has not been clear—until now.

The P&G team discovered a novel correlation between the presence of spots and an increase in the dendricity of melanocytes through comparing spot area versus nonspot area in multiple types of facial spots, including melasma, solar lentigo, and postinflammatory hyperpigmentation (PIH)/acne marks. The findings, which indicate that "increasing the export network of melanosomes may be a key common mechanism for increased pigmentation in spots," were reported in their groundbreaking paper for the *Journal of the European Academy of Dermatology and Venereology*. In essence,

Tomohiro Hakozaki

PHOTO: COURTESY OF PROCTER AND GAMBLE

the greater the dendricity, the more channels there are for melanin to be deposited on the surface of the skin. And the team uncovered another fascinating correlation: a link between the proinflammatory protein high mobility group box 1 (HMGB1) and "skin pigmentation machinery." Indeed, this pesky protein is a key contributor to stimulating formation of melanocyte dendricity.



Immunofluorescence and cell culture studies from Procter & Gamble scientists show that increased melanocyte dendricity is linked to hyperpigmented skin spots and that a combination of sucrose dilaurate and sucrose laurate suppresses melanocyte dendricity in vitro.

PHOTO: ADAPTED FROM J. WANG *ET AL*. THE COMBINATION OF SUCROSE DILAURATE AND SUCROSE LAURATE SUPPRESSES HMGB1: AN ENHANCER OF MELANOCYTE DENDRICITY AND MELANOSOME TRANSFER TO KERATINOCYTES. *JEADV* 2022, 36 (SUPPL. 3), 3-11.

With this discovery, the P&G team aimed to examine what compound could potentially suppress the development of this dendricity in vitro. In collaboration with a supplier, they identified a special ratio of sucrose dilaurate (SD) and sucrose laurate (SL) in combination, called SDL; these two compounds have been used for years as emulsifiers, dispersants, or stabilizers in cosmetics, foods, and pharmaceuticals, with no safety issues. However, the team demonstrated that the SDL combination produced a "stabilizing" effect—a marked reduction of HMGB1 release, and a significant decrease in melanocyte dendricity formation and a reduction of dendrites already formed. "SDL can effectively stabilize the melanocyte from the root, which is a breakthrough in vitro science discovery in the industry regarding the spot-fading mechanism," Hakozaki notes. Moreover, SDL also reduces the melanosome transportation machinery protein Rab27a through the dendrite tips within melanocytes in a coordinated manner, inhibiting melanin production and transfer and lowering the total melanin amount in the epidermis.

"To prove this, we did a lot of sequential in vitro experiments," Hakozaki says. "There are many ways to reduce melanin production. Most of them relate to melanin synthesis. This is the only one we found so far to modulate the melanocyte dendricity effectively. So [for me] it was a 'holy cow' moment as a team member!"

This advancement has the potential to create whole new branches of technology. "Human biology is so complex," Hakozaki says. "We believed there should be much more, some other mechanism involved here." And now that they know what that mechanism is, scientists can develop new technology for targeting spots. For example, P&G has been adding SDL to niacinamide (a.k.a. vitamin B3), an invention that has recently been granted a U.S. patent.

It makes sense that P&G scientists would be the ones who unraveled the science behind spots. The company has always invested heavily in R&D and has pursued a science-first approach in improving people's lives.

There are still abundant opportunities to explore. "A spot is a very small place on your skin, but it involves very complex biology and pathways. What we have achieved is discovering a new intervention pathway in vitro," says Hakozaki. "But there is more to discover that can help the whole skincare industry in scientific research." He continues to be engaged and energized by this research. "I love my job," he says. "There is a saying that 'the only way to do great work is to love what you do.' If you haven't found it, keep looking. Don't settle. As with all matters of the heart, you'll know it when you find it. This is exactly what I have [found] here. It is really exciting to be working on cutting-edge science with thousands of talented people in specialties such as biology and beyond." Clearly, P&G values Hakozaki, because as he says, "the company has allowed me to stay focused in this field for more than a decade." And for an innately curious scientist, dedicated to improving the lives of people around the world, that hits the spot.

Sponsored by

OLAY

CONTENTS

18 MARCH 2022
VOLUME 375
ISSUE 6586

NEWS

IN BRIEF

IN DEPTH

FEATURES

INSIGHTS

POLICY FORUM

PERSPECTIVES

BOOKS *ET AL.*

PHOTO: CORNELL WATSON

LETTERS

RESEARCH

IN BRIEF

REVIEW

RESEARCH ARTICLES

REPORTS

A multiseason field experiment shows that fruit flies can quickly adapt to environmental change.

PHOTO: ERIC SUCAR/UNIVERSITY OF PENNSYLVANIA

DEPARTMENTS

ON THE COVER

White clover (*Trifolium repens* L.) populations in Los Angeles, such as the one shown here, evolve to have less hydrogen cyanide, a defensive chemical that also affects tolerance to abiotic stressors. A new study reveals that these plants repeatedly adapt to urban environments throughout the world, indicating that cities are drivers of evolutionary change across the planet. See page 1275. *Photo: Ashley Crichton*

SCIENCE (ISSN 0036-8075) is published weekly on Friday, 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 © 2022 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): $2212; Foreign postage extra: Air assist delivery: $98. 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.

EDITORIAL

A fabric of life view of the world

Sandra Díaz
is a professor and senior investigator at the Instituto Multidisciplinario de Biología Vegetal (IMBIV), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), and Universidad Nacional de Córdoba, Córdoba, Argentina.
sandra.diaz@unc.edu.ar

This month in Geneva, the 196 parties to the United Nations (UN) Convention on Biological Diversity will discuss the post-2020 Global Biodiversity Framework, in preparation for part two of the UN Biodiversity Conference (COP 15) in Kunming, China. By driving biodiversity actions worldwide through 2030 and beyond, this is arguably the most important biodiversity policy process of our time.

There is a general sense that time is running out. Policy-makers and the general public are increasingly well-informed about nature, but this has not translated into slowing down its fast deterioration. Most of the biodiversity goals that have gained public attention are not consistent with the connectedness between humans and other organisms, or between different places and peoples through living bodies. This is in part reinforced by technical definitions, metaphors, and social narratives associated with biodiversity that are not explicit about this connectedness. Public understanding is crucial because to achieve the transformative change called for by nearly all recent environmental reports, all sectors of society—not just academics and policy-makers—must weigh in on decisions about the future of life on Earth. Better ways to convey this picture are needed.

> "...a metaphor that better resonates with...society may spark... meaningful action."

The expression "fabric of life" has long been used in lay communication to refer to the whole of the living world. The fabric of life on Earth has been described as "woven" by natural processes over many millions of years and in conjunction with people—our livelihoods, our institutions, our stories—for many thousands of years. The expression is starting to emerge at the interface of science and policy, such as in the 2019 Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services Global Assessment Report. Thus, far from being a quaint figure of speech, this expression could be a useful, timely, actionable metaphor to facilitate transformative change.

Why dwell on metaphors at a time when each day of inaction counts? Because action is urgently needed. Metaphors help make sense of the world. They provide a scaffold for thinking, which in turn frames action.

The fabric of life framing shines a spotlight on our deep evolutionary kinship with, and physical and cultural dependence on, the rest of the living world. Although not intended to replace the term "biodiversity" in ecological research, a metaphor that better resonates with other disciplines and sectors of society may spark new questions, engage the public imagination in inspirational ways, and mobilize meaningful action. I have seen people who would normally stare politely but without interest at a presentation about the state of biodiversity suddenly listen with engrossed attention when hearing about the fabric of life, humanity's entanglement within it, and the vast richness of nature's contributions to people.

In the context of the upcoming UN meetings, the fabric of life framing could help design a more powerful set of goals, targets, actions, and indicators in intergovernmental biodiversity strategies. The evidence is overwhelming for pervasive and deep connections between people and the rest of the living world everywhere, and for socioecological systems increasingly being influenced by distant demands and decisions. However, the science-policy interface has yet to catch up. For example, the first segment of the UN Biodiversity Conference in October 2021 saw a strong emphasis on goals, targets, and pledges that deal with the immediate, local causes of biodiversity decline. This focus could be broadened to tackle the systemic and connected underlying causes that are embedded in the ways people consume, trade, and allocate subsidies, incentives, and safeguards. This could take, for example, the form of redeploying current subsidies, incentives, and penalties toward activities that preserve and ideally help "reweave" the fabric of life. Another example is the present emphasis on expanding protected areas. Equal attention should be paid to shared landscapes and seascapes that make up the vast proportion of the planet's surface and where people's everyday entanglement with other organisms is the most obvious.

By thinking about the living world as an intermeshed fabric, we start to shift (or broaden) the spotlight of inquiry and action, making them more focused on connections and entanglement, and more interdisciplinary and socially inclusive. This is a small but necessary step to rally a wider range of society into producing new knowledge and spurring action for a better future. Suitable metaphors have been instrumental to many social changes in the past; we should not underestimate their importance this time.

–Sandra Díaz

PHOTO: DANIEL M. CÁCERES

Killing the golden goose for STEM

In 2016, the US Congress enacted the Puerto Rico Oversight, Management, and Economic Stability Act (PROMESA), which led to a federally appointed Financial Oversight and Management Board (FOMB) to oversee restructuring of the island's $74 billion debt and achieve sustainable budgets. Many economists argue that even debt adjustment, which was approved in January, will not be enough to put Puerto Rico on a road to recovery. Unprecedented budget cuts by FOMB have already reduced funding from Puerto Rico's state government for the University of Puerto Rico (UPR). UPR is the island's premier institute of higher learning in science, technology, engineering, and mathematics (STEM). State funding for UPR must be restored if Puerto Rico is to stably recover from years of economic despair.

In knowledge-based economies, investments in academic programs facilitate research and development (R&D) and power economic growth. STEM fields generate two-thirds of jobs in the United States, 69% of the US gross domestic product, and over $2 trillion in annual federal tax revenue. Research universities play a major role in generating opportunities for R&D investment and in training the STEM workforce to meet the needs of a vibrant economy. This is especially true in Puerto Rico. UPR is the archipelago's largest public higher education system, with 11 campuses and 694 academic programs covering a broad range of disciplines. Since 2017, state budget support for UPR has been reduced by 48%, severely affecting STEM research and education programs. UPR oversees 80% of higher education funding for R&D in Puerto Rico and supports 79 research centers across the territory. Its Río Piedras campus is the only Puerto Rican institution certified as an R2 Doctoral Research University by the Carnegie Classification, a designation that applies to only 133 of the nearly 3936 colleges and universities across the United States. UPR is also a leader in scholarly production across STEM disciplines and among all Caribbean institutions. In addition, the university is a major contributor to local economies—every dollar invested in the UPR generates $1.56 of economic activity in other sectors.

Less recognized is the role that UPR campuses have in STEM training of Hispanics and other minority populations in the US. At the national level, data from the National Science Foundation show that from 2015 to 2019, when considering the top 50 US baccalaureate institutions of Hispanic or Latino STEM doctorate recipients, 16.8% of students received a bachelor's degree from UPR, more than any other institution in the nation. When considering the top 14 baccalaureate institutions of Hispanic female and Latina STEM doctorate recipients, the largest numbers of such graduates came from four UPR campuses. The UPR Mayagüez campus ranks second among US institutions that award bachelor's degrees in engineering to Hispanic students. For Hispanic or Latino students that were awarded a PhD in engineering, the majority had bachelor's degrees from UPR Mayagüez. It makes socioeconomic sense to restore investment in UPR given its role in reducing the ethnicity and gender gaps in the US STEM workforce.

> “...budget cuts... damage the pipeline of... Hispanic and Latino students into the STEM workforce.”

Since 2017, draconian budget cuts and austerity measures imposed by the FOMB have threatened the inclusive scientific environment and training that UPR provides for permanent residents in the US and Puerto Rico. UPR has suffered a 23.4% reduction in personnel in less than 4 years. This is not only demoralizing but also jeopardizes the university's capacity to drive economic R&D and produce STEM professionals for the US and Puerto Rico.

The focus of the FOMB's investment strategies should be on building UPR's capacity, rather than contributing to the demise of this golden goose for STEM. On average, US states and local governments target 9.4% of their general spending for higher education. Puerto Rico used to spend around 9.6%, but with budget cuts it is now spending close to 6%. Such reductions not only hurt Puerto Rico but also damage the pipeline of highly qualified US Hispanic and Latino students into the STEM workforce.

In 2020, Congressman Raúl Grijalva proposed an amendment to PROMESA (H.R.6975) that provides the UPR with a minimum budget that approaches the 2017 amount before the austerity cuts. The bill has not yet reached Congress. It is time for the White House to evaluate the effects of budget cuts on UPR's STEM research and education. PROMESA needs to fulfill its promise to Puerto Rico. A battered UPR will result in a weakened and less diverse US STEM workforce.

–Elvia Meléndez-Ackerman and Jorge L. Colón

Elvia Meléndez-Ackerman is a professor in the Department of Environmental Science at the University of Puerto Rico–Río Piedras, San Juan, Puerto Rico. elvia.melendez1@upr.edu

Jorge L. Colón is a professor in the Department of Chemistry at the University of Puerto Rico–Río Piedras, San Juan, Puerto Rico. jorge.colon10@upr.edu

10.1126/science.abo1138

NEWS

IN BRIEF

Edited by **Jeffrey Brainard**

Sea creatures colonized the *Endurance*'s rear rail and helm wheel on its aft well deck.

EXPLORATION

Endurance preserved in Antarctic depths

More than 100 years after it was crushed by ice and sank in the Weddell Sea, the *Endurance*, the famed ship of Antarctic explorer Ernest Shackleton, was discovered last week in near pristine condition. Searching with an underwater robot for 2 weeks amid sea ice east of the Antarctic Peninsula, the Endurance22 team, funded by an anonymous donor, finally spotted the 44-meter ship sitting upright on the 3000-meter-deep sea floor. It was found only 6 kilometers from its last known location. A crew of anemones, sea squirts, and sponges has taken charge of the ship. In the frigid depths, microbes spared the ship's wood; broken masts and the helm wheel were preserved. The team will not touch the ship, which is protected by the Antarctic Treaty, but will scan its remains. Despite Shackleton's failure to land on Antarctica, let alone cross it, he has remained a revered figure in polar exploration for his leadership in saving his entire crew following the destruction of his ship in 1915.

PHOTO: FALKLANDS MARITIME HERITAGE TRUST/*NATIONAL GEOGRAPHIC*

East Asia sees COVID-19 explode

PANDEMIC | China's "zero COVID" policy is facing its stiffest challenge after a rapid rise in cases caused by the Omicron variant that has forced authorities to take draconian control measures. Domestically acquired COVID-19 infections jumped from 175 on 7 March to 3507 on 14 March. Asymptomatic infections, which China tracks separately, surged as well. The government has put 37 million people in the southern city of Shenzhen and the northeastern province of Jilin under lockdown orders, with many elsewhere facing travel and other restrictions. Public health experts worry the increases might overwhelm the country's heath system, as it already has in Hong Kong. South Korea and Vietnam, which managed to keep cases very low during the first 2 years of the pandemic, have also seen explosions due to Omicron, with South Korea now reporting more than 300,000 cases daily and Vietnam more than 200,000. Many European countries that recently relaxed or abandoned control measures are recording COVID-19 rebounds as well, with double-digit percentage increases in infections over the past 2 weeks.

Pig heart recipient dies

BIOTECHNOLOGY | David Bennett, who in January became first person to receive a heart transplant from a genetically engineered pig, died on 8 March, physicians from the University of Maryland School of Medicine announced last week. Bennett, who had advanced heart failure and was ineligible for a human heart transplant, received a heart from a pig created by the biotech company Revivicor. The organ had been genetically engineered to reduce rejection by the human immune system and keep the heart from outgrowing a person's chest. No cause of death was immediately released, but the Maryland team that performed the transplant said in a statement that the organ "performed very well for several weeks without any signs of rejection."

DISASTERS

Australian universities endure climate change's toll

Record rainfall on Australia's east coast last week destroyed dozens of research projects in greenhouses at the University of Queensland (UQ)—a fresh reminder of the risks posed by climate change to the country's academic facilities. Some areas received nearly 1 meter of rain in a week, prompting the country to declare its first national state of emergency. UQ's greenhouses (above) also flooded in 2011; they were rebuilt at the same site because administrators believed the flood was a once-in-a-century event, UQ plant biotechnologist Ian Godwin says. This time, university administrators have indicated the greenhouses will be moved to higher ground or built on stilts. In December 2019 and January 2020, universities in Sydney and Canberra canceled classes after buildings were inundated with bushfire smoke.

Pandemic deaths undercounted

PUBLIC HEALTH | COVID-19 took more than 18 million lives by the end of 2021, roughly three times as many as the officially reported toll, according to a controversial new estimate. Several earlier studies indicated that tallies by the World Health Organization (WHO) vastly underestimate deaths, but the new one, published in *The Lancet* last week, is the first peer-reviewed examination of the period through December 2021. Researchers from the Institute for Health Metrics and Evaluation (IHME) analyzed figures from 74 countries and territories to estimate excess mortality—deaths exceeding those expected from other causes. India's official numbers omit 3.5 million COVID-19 deaths, the largest gap between reported and estimated ones. But Egypt and several other countries in Africa with smaller populations have bigger ratios of estimated to reported deaths. The IHME analysis has been criticized for overestimating excess COVID-19 deaths, particularly in some higher-income countries. WHO plans to release its own revised estimates in the next few weeks.

Death disparity

Estimated excess mortality from COVID-19 in this regionally diverse sample of 10 countries accounts for nearly half the new estimate of total excess deaths attributable to the disease.

COUNTRY	REPORTED	ESTIMATED EXCESS	EXCESS/ REPORTED
India	489,000	4,070,000	8.33
USA	824,000	1,130,000	1.37
Russia	651,000	1,070,000	1.64
Mexico	418,00	798,000	1.91
Egypt	21,800	265,000	12.19
Italy	137,000	259,000	1.89
Ethiopia	6,940	208,000	30
Japan	18,400	111,000	6.02
Uzbekistan	1,480	69,100	46.51
China	4,820	17,900	3.71
Global total	**5,940,000**	**18,200,000**	**3.07**

CREDITS: (PHOTO) LEE HICKEY; (DATA) H. WANG *ET AL.*, *THE LANCET*, HTTPS://DOI.ORG/10.1016/S0140-6736(21)02796-3 (2022)

arXiv pulls 'inflammatory' papers

PUBLISHING | Administrators of arXiv, a preprint server used by many physicists, recently removed or refused to post several papers from opposing sides of a debate over claims of room temperature superconductivity, saying their manuscripts include inflammatory content and unprofessional language. Last month, arXiv also banned one of the authors, Jorge Hirsch, a theoretical physicist at the University of California, San Diego, from posting papers for 6 months. He calls the ban "very unfair"; some other physicists accused arXiv of squelching science. But the server's administrators argue the decision wasn't about the merits of the research. Volunteer moderators reject only about 1% of arXiv's estimated 15,000 monthly submissions. Launched in 1989, arXiv has ballooned to host more than 2 million preprints and has become a vital communication hub for physicists and other scientists to discuss new findings before they are submitted to peer-reviewed journals.

Sweden's pandemic chief leaves

PUBLIC HEALTH | The controversial chief epidemiologist at Sweden's Public Health Agency has stepped down to become a consultant for the World Health Organization (WHO), focused on pandemic vaccinations. Anders Tegnell, the unflappable voice of Sweden's mostly laissez-faire approach to COVID-19, has many fans inside and outside the country—including a few who had his face tattooed on their arm or leg. But the response he led was criticized by two official commissions in Sweden, particularly because of the country's high death toll early in the pandemic. On 9 March, the health agency announced that Tegnell would leave to join a committee that will help coordinate vaccination efforts by WHO, UNICEF, and

THREE QS

Sequencing Africa's richness

To help protect his homeland's biodiversity and aid its agriculture, Nigerian biochemist ThankGod Echezona Ebenezer and colleagues in 2021 founded the Africa BioGenome Project (AfricaBP). Its broader goal: sequence every plant, animal, and other eukaryote native to the continent, estimated at more than 100,000 species. The effort now includes 109 African scientists and 22 African organizations, representing all five regions in the African Union. In a commentary this week in *Nature,* Ebenezer calls on donors and governments, in Africa and beyond, to commit $1 billion over the next decade to AfricaBP. (A longer version of the interview is at https://scim.ag/AfricaBP.)

Q: Why do this project?

A: When I started my Ph.D., I searched databases for bioinformatics and genomes of plants and animals in Africa—not much was there. I was interested in how fishes adapt to environmental conditions through the help of parasites. I came across *Bostrychus africanus*, a fish mainly found in West Africa that is host to a nematode parasite, *Raphidascaroides africanus*. I checked sequence databases, and limited information existed. It became apparent to me that African species are neglected by the global genomic community.

Q: How do you plan to do it?

A: We aim to ensure that at least 80% or more of the sequencing happens on the continent by African scientists through our program. Even if they are going to be supported by scientists from the global north, they should play a leading role because it is only by doing so that they will be able to do it on their own in the future. … For the Human Genome Project, there was no Africa involvement—what this meant is the continent lost out.

Q: What are some of the challenges?

A: At the moment, few African institutions have positions for bioinformatics and genomicists. Where they do, they are not permanent. … [But] I see the future as bright. I have had so much interest from students from the continent. There is a huge buy-in from early career scientists to established scientists and governments officials.

SCIENCE.ORG/NEWS
Read more news from *Science* online.

Gavi, the Vaccine Alliance, a public-private partnership that supports inoculations in low-income countries.

Plagiarism fells research chief

SCIENTIFIC MISCONDUCT | The vice chancellor for research at the University of North Carolina, Chapel Hill, resigned that position last week after a federal regulator found he included plagiarized material in a grant application for cancer research. Terry Magnuson had held the post, which oversees investigations of research misconduct on the campus, since 2016. The plagiarized material involved descriptions of methodology and other content from online corporate guides and a published paper. In a statement to colleagues, Magnuson explained that the "mistake" resulted because he placed the text in the proposal as a placeholder and, distracted by his administrative duties, failed to revise or credit it. "I hope this experience serves as teachable moment for us all," he wrote. He will remain a professor of genetics, and his work funded by the U.S. Public Health Service will be subject to supervision until January 2024.

New home for tropical gene bank

AGRICULTURE | The world's largest repository for seeds and tissues of beans, cassava, and tropical forage grasses has a new home. This week, the International Center for Tropical Agriculture and Bioversity International dedicated a $17 million facility, Future Seeds, in Palmira, Colombia, with a new genomics lab and 30% more storage than its predecessor there. The collection currently contains nearly 67,000 specimens. Plant breeders in many countries have used them to create better performing and climate resistant crops.

A U.S. Coast Guard research team recovers a supply drop amid melted Arctic sea ice in July 2011.

PHOTO: NASA/KATHRYN HANSEN

CLIMATE CHANGE

Multiyear sea ice vanishing rapidly

In another sign of global warming's effect, the Arctic has lost one-third of its winter sea ice volume in the past 20 years, including a large amount of the thick, "multiyear" ice that survives summer melting, a study has found. Although the declining extent of sea ice is easily observed, measuring its thickness is more difficult and requires probing from space by lasers or radar. The new work combined a 3-year record of ice thickness from NASA's ICESat-2 satellite with older satellite measures; the decline in volume was driven especially by a replacement of multiyear ice by seasonal ice, which forms in winter and melts the following summer, researchers reported last week in *Geophysical Research Letters*. Climate change is warming the Arctic four times faster than the rest of the world, and most climate scientists expect the region to be free of summer ice by 2050, if not sooner.

Ukrainian civilians evacuate from the embattled town of Irpin, a suburb of Kyiv.

PHOTO: GIANLUCA CECERE/LAIF/REDUX

EUROPE

Ukrainian researchers flee war trauma and terror

Refugees find aid and jobs, but many remain to fight Russian invaders

By **Richard Stone**

At this time of year Maryna Kravchenko, a population ecologist at Kharkiv National University (KNU), normally heads into the forests of eastern Ukraine to track frogs and toads emerging from hibernation. Instead, after 1 week sheltering from bombs in a basement, she fled with her young children to Germany, where colleagues at Ludwig Maximilian University of Munich had found them an apartment. Her husband, KNU zoologist Dmitry Shabanov, stayed to defend Kharkiv. "We've all stopped wishing each other good morning," Kravchenko says. "Instead, it's now 'How are you?' to both check whether the other person is still alive and show that you still care about them."

The humanitarian toll of an increasingly brutal war is rising. More than 3 million refugees have fled since Russia began its invasion on 24 February, according to the Office of the United Nations High Commissioner for Refugees. Scientists like Kravchenko are among the dispossessed. Others are staying put, hunkering down or taking up arms in the war zone. The upheaval is touching the Russian science community as well. Cut off from many international collaborations and, in some cases, opposed to the war, some Russian scientists are fleeing their own country.

Before the invasion, Ukraine had nearly 80,000 scientists, says Brokoslaw Laschowski, a Ukranian-Canadian postdoctoral fellow at the University of Toronto. He's volunteering for an impromptu international effort, Science for Ukraine, that has lined up hundreds of jobs for refugee scientists—mainly women, as Ukraine forbids most men under 60 from leaving the country. When war broke out, "My first thought was that one generation or two generations of scientists were going to be destroyed," says Sergio Ponsá Salas, a chemical engineer at the University of Vic-Central University of Catalonia who's helping coordinate Science for Ukraine in Spain. His university has already thrown a lifeline to three food scientists and a hydrologist who have escaped.

Others want to stay. The day after fighting began, KNU astronomer Oleksiy Golubov, 36, rushed to the local office of the Territorial Defense Forces. A recruiter took one look at his left hand—limp from a hereditary nerve disorder—and turned him away. "He told me I'm not fit. That I wouldn't be able to hold a weapon properly," says Golubov, his voice hoarse with emotion.

On 7 March, Golubov and a group of fellow scientists escaped the besieged city, only half an hour's drive from the Russian border. He's now holed up at the home of a family friend in Batkiv, a village in western Ukraine. There he pulled off a second escape—this time, an intellectual one—when he joined a virtual poster session last week at the Lunar and Planetary Science Conference in Houston to present research about an iron meteorite recovered in 2021 in Sweden. When attendees learned about his plight, he says, "it became impossible to speak about science." On the spot, several European colleagues offered to host him in their labs. "I'm grateful," Golubov says, but he doesn't intend to leave Ukraine.

As the Russian onslaught intensifies, civilian casualties are mounting and infrastructure is taking a pounding, including scientific facilities. One day after a deadly missile strike on a maternity hospital on 9 March in Mariupol, shells fell on the city's Pre-Azov State Technical University and on the Kharkiv Institute of Physics and Technology, damaging a linear accelerator and a subcritical nuclear reactor that generates neutrons for experiments. That same day, Russian forces overran and destroyed the Institute of Irrigated Agriculture in Kherson, its director told local media.

Several days earlier, a rocket strike on Kharkiv's city administration building blew out the windows of a KNU building on the other side of Freedom Square. "Broken glass everywhere and not a soul," Vadym Kaydash, director of KNU's Institute of Astronomy, emailed staff after visiting the scene the next day. When the missile hit, says Glib Mazepa, who lives three blocks from the square, "Our building started swaying like a ship in a storm." A graduate student at the University of Lausanne and Uppsala University, Mazepa returned to his native Kharkiv last year to save money, he says, as he finishes a Ph.D. on frog evolution.

As the Russian bombing of Kharkiv grinds on, Mazepa is working with friends to secure collections in the city's museums and universities, one of a number of frantic efforts nationwide to safeguard specimens and research samples (see story, p. 1210). Mazepa and his wife have also learned a grim new skill that's in high demand in Kharkiv these days: making Molotov cocktails.

> **"My city may be ruined, but we will build it back."**
>
> **Maryna Kravchenko**, Kharkiv National University

Veronika Lipatova, a KNU astronomer and member of Golubov's asteroid modeling group, initially served as a volunteer medic, sourcing medicines and triaging wounded soldiers while her colleagues moved computers and astronomical cameras into a cellar that was used to safeguard equipment when Nazis occupied Kharkiv during World War II. Last week, Lipatova evacuated with Golubov to western Ukraine, but keenly regrets that decision. "I'm here, safe. My friends in Kharkiv aren't. I feel bad," she says, fighting back tears. "Like I'm not a good human for leaving."

Some Russian scientists are ashamed of the war their nation is waging. "It's a very bad feeling when you can't do anything to stop it," says Ilya Schurov, a mathematician with the Higher School of Economics University in Moscow. "It's a moment in which you find yourself in hell." He and his wife joined a street protest on 27 February that he had publicized on social media. He also signed two letters denouncing the war. After learning the police were looking for him, the couple found seats on a flight to Tajikistan, and then flew on to Istanbul. "My professional life is ruined," Schurov says. But he doesn't regret taking a stand—nor does he plan to return to Russia anytime soon.

With President Vladimir Putin's increasingly totalitarian regime cracking down on protestors, many Russian scientists may be too frightened to speak out. KNU astronomer Irina Belskaya, who knows many Russian scientists from conferences and past collaborations, says she heard from just one, in Moscow, who congratulated her on International Women's Day on 8 March. "I replied, 'Do you realize now there's a war? Russia is bombing Kharkiv University.' He wrote back simply that he did not vote for Putin."

Others may be oblivious to the true nature of the war, with Russian state media pushing propaganda and disinformation. The National Research Foundation of Ukraine says it emailed 49,000 Russian scientists with an appeal to speak out against the war. The vast majority did not respond; many who did sent responses the foundation claimed were "obscene, full of imperial malice and chauvinism." Mazepa says he got a similar reaction when he emailed Russian friends and colleagues from joint expeditions and projects. "These people don't respond, or [they] say, 'Bullshit, you are not being bombed, it's a liberation,'" Mazepa says. "I always imagined it would be the role of scientists and artists to speak out, but to me it looks like the majority of Russian scientists are supporting this nightmare."

Although he has not taken up arms, Golubov is engaged in the fight. After wrapping up his conference presentation, he got back to posting war updates on Facebook and reassuring worried friends that he's still alive. (KNU's Facebook page notes that two of its professors and a student have been killed.) He's particularly proud of some computer jocks in his institute who he claims hacked into a Russian state TV broadcast to show brief video clips of atrocities committed against Ukraine—and dead Russian soldiers. "At the moment, this is more important than creating computer programs for simulating astronomical objects," he says.

The war's scars are likely to linger. After 1 week huddling in a bunker, Volodymyr Nemchenko, a 73-year-old computer scientist from Kharkiv National University of Radio Electronics, escaped with his wife to Luxembourg, where his son lives near an airport. "Every time a plane comes near, we're afraid," Nemchenko says. When they hear a loud noise, like a door slamming, "we're shocked," he says.

The outcome of the war is far from certain. But Kravchenko, for one, knows what she will do if Ukraine prevails. "I'm crystal clear that I want to go home," she says. "My city may be ruined, but we will build it back." Golubov too is optimistic. "Ukrainians are more united than ever," he says. "When the war ends, and we're able to return to science, it will be the best science we've done in our lives." ■

With reporting by Andrew Curry and Rodrigo Pérez Ortega

EUROPE

Ukrainians rush to save 'irreplaceable' collections

Researchers hide ancient weapons and bat skulls, upload data as bombs fall

By **Andrew Curry**

On the morning of 24 February, conservation biologist Anton Vlaschenko awoke to the sound of shelling outside his apartment in Kharkiv, Ukraine. The first thing he did was eat a big breakfast. Then, he headed straight to the Ukrainian Bat Rehabilitation Center, the largest bat rescue and research facility in Eastern Europe. "I didn't know if we would return home, or what would happen next," he says. "But I understood the war had begun, and we needed to do something."

Worried that the city's power would be cut, Vlaschenko spent the next 24 hours transferring hundreds of rescued bats hibernating in the center's refrigerators to special cages for release. As the winged mammals flew into the frigid night to look for new spots to spend the remaining winter, Vlaschenko heard gunfire: the first Russian troops to enter the city outskirts were clashing with the Ukrainian army.

Next, Vlaschenko moved the center's collection of more than 2000 *Nyctalus noctula* bat skulls—each carefully padded with shredded newspaper and labeled and stored in a numbered matchbox—to his apartment, an hour's walk away. More than a week later, the skulls were still there, wrapped in plastic shopping bags by the door in case they need to be moved again in a hurry. He also brought home rescued bats too sick to release.

"We had a huge explosion close to my home 2 days ago," Vlaschenko said in a call last week from his apartment, the peeps of bats audible in the background. "You never know what moment you could be hit."

As war rages, Vlaschenko and researchers across Ukraine are scrambling to protect, hide, or evacuate irreplaceable specimens, collections, and data, and uploading 3D scans of specimens to colleagues abroad. For heritage experts, the threat to Ukraine's

Anton Vlaschenko holds a rescued bat at his home in Kharkiv while news coverage plays on his cellphone.

scientific collections and cultural monuments is frighteningly familiar from recent conflicts in Iraq, Syria, Mali, and elsewhere. “How do you protect museums? You can’t move the buildings, or the infrastructure,” UNESCO World Heritage Centre Director Lazare Eloundou Assomo says. “You try to protect collections by relocating them to shelters or refuges, where you have to hide and store them until the war is finished.”

But not all collections can be easily moved. With colleagues, Kharkiv National University herpetologist Oleksandr Zinenko helps curate 20,000 birds and 8000 amphibians and reptiles collected over the past century, many stored with countless invertebrate specimens in wood and glass cases lining the arched halls of the university’s natural history museum. “The collection is a time machine to look at the changing conditions of the past. It’s irreplaceable,” says Zinenko, who is now in Lviv. “It needs to be protected, but you cannot just put it in your pocket and go away. How do you choose what to save?”

Colleagues still on the ground in Kharkiv says the museum building, built in 1905, is still standing, but bomb strikes blew out many of its windows, and temperatures have plummeted to –10°C as snow blankets the ground outside. “Some employees [are] living in the museum and guarding the collections right now,” Zinenko says. “Part of me feels I should be there too.”

In a blog post days after the Russian invasion began, National Museum of the History of Ukraine Director Fedir Androshchuk said his museum in central Kyiv had dismantled exhibitions with artifacts such as Scythian weapons and a mammoth tusk bracelet from the last ice age and moved them to secure spots. Androshchuk fears “eventual damages caused by rocket strikes, shelling and bombing,” he wrote in an email to *Science* last week. “There is no guarantee that Ukrainian heritage will be safe.”

In the early days of the war, when some predicted Kyiv would fall quickly, it wasn’t clear how much time museum staff had. “They were working against the clock,” says Mads Holst, director of the Moesgaard Museum in Aarhus, Denmark, who has been in touch with Androshchuk since the war began.

A plinth from the 10th century C.E., now on exhibit in Denmark, bears the trident emblem used as a Ukrainian national symbol today.

In what Holst calls a “happy accident,” more than 1000 objects from the National Museum of the History of Ukraine and regional museums had already been shipped to his museum for an exhibit, *Rus—Vikings in the East*. “There are burial assemblages, hoard finds—they’re quite significant objects,” Holst says. “There’s still quite a lot to be explored and researched about [them.]”

Many Ukrainians view the period between 800 and 1050 C.E., when Kyiv was founded as a Viking trading outpost, as the beginning of their national identity—an interpretation many Russians, including President Vladimir Putin, deny. Now secure and on display in Denmark, “these objects are part of this conflict, which is partly about whether Ukraine is allowed to have an independent historical identity,” Holst says.

Other researchers are turning to fragile internet connections to save their collections digitally. Pavel Gol’din, an evolutionary zoologist at the Schmalhausen Institute of Zoology in Kyiv, is far from the front lines—for now—in the southwestern Ukrainian city of Chernivtsi. His work relies in part on massive cetacean fossils 10 meters long or more, which would be nearly impossible to move.

Over the past 2 years, he led a project to scan marine mammal fossils from collections housed across Ukraine. “We created a 3D archive of extinct and extant specimens, some unique, … [including] four terabytes of data—it’s quite a big collection.”

When war broke out, the scans were mostly on hard drives in Kyiv and Kharkiv. Last week, to make sure at least the scans survive, Gol’din’s graduate student Pavlo Otriazhy, then in Kharkiv, transferred data to colleagues in France in the midst of Russian bombardments, a gigabyte every 10 minutes or so. When his internet connection became too unstable, Otriazhy took regional trains to reach Dnipro, 200 kilometers to the south, hard disks in his bag, and kept uploading. “If something happens to me, other scientists can still access [the scans]. That’s much safer than on a hard disk.”

Gol’din and Otriazhy are working to help avert a repeat of the catastrophes that befell paleontology during World War II, when bombs and fire destroyed countless fossils in museums from Munich to Milan and Kyiv. “I think we will win,” Gol’din says. “We will be able to start again.”

Back in Kharkiv, Vlaschenko’s bat center has lost just a few windows and its power is still on. The center’s veterinarian and a few volunteers have stayed in the city to care for bats too sick to release. Vlaschenko says the bat collections in his apartment and at the center represent decades of work—and the hope that Ukrainian science can recover after the war.

“When something like this happens the specimens that survive are crazy important for future research,” Vlaschenko says. “You can buy new equipment or build new buildings, but you can’t get back individual specimens.” ■

PHOTOS: (TOP TO BOTTOM) ANTON VLASCHENKO; MOESGAARD MUSEUM

A U.S. Army paratrooper trains a Polish soldier after Russia's invasion of Ukraine.

PHOTO: ARMY SGT. CLAUDIA NIX/U.S. DEPARTMENT OF DEFENSE

U.S. BUDGET

Growth at U.S. science agencies curtailed in final budget deal

Congress boosts defense, shrinks Biden's research vision

By **Jeffrey Mervis**

Last week, Congress agreed to give roughly 5% more money to many federal science agencies when it passed a $1.5-trillion government budget for 2022. But what science advocates had hoped would be a banner funding year fell victim to a bipartisan push to spend much more on defense than President Joe Biden had requested—and to one Democratic senator's rejection of a separate Biden plan to allocate billions of additional dollars for research.

Besides beefing up the U.S. military, appropriators reduced by more than half the 16%, $110-billion increase Biden had sought for civilian programs—including a slew of new and expanded research initiatives. Biden had proposed boosting defense programs by only 1.8%, or $12 billion, but the annual spending package he signed into law on 15 March gives defense a 5.6% boost and civilian programs a 6.7% bump. Having a final 2022 budget also ends a 5-month governmentwide spending freeze that prohibited any new initiatives.

For the National Institutes of Health (NIH), the government's biggest funder of academic research, that shift in priorities shrunk Biden's proposed 21% hike, much of it for a new Advanced Research Projects Agency for Health (ARPA-H) to develop cutting-edge medical treatments, to 5%. The final bill, which gives NIH $45 billion, means ARPA-H will be launched with $1 billion rather than the $6.5 billion Biden had requested (see story, p. 1213).

The National Science Foundation (NSF) had an even wilder roller-coaster ride. In addition to Biden's request for a 20% boost, to $10.2 billion, NSF was slated to get several billion more under the president's proposed $2.2 trillion social welfare initiative, called Build Back Better (BBB). Some of the money was for a new research directorate aimed at accelerating emerging technologies. But Senator Joe Manchin (D–WV) balked at BBB's scope and price, dooming efforts to pass the legislation in the evenly divided Senate. NSF ultimately received $8.84 billion, a 4% increase—but no additional funds to launch its Technology, Innovation, and Partnership (TIP) directorate.

The demise of Biden's plan also deprived the Department of Energy's (DOE's) Office of Science of several billion dollars for building and upgrading user facilities at its 10 national laboratories. Even so, Congress gave it a 6.4% boost, to $7.48 billion. Applied energy programs aimed at combating climate change did better, jumping 12% to $3.2 billion. NASA's science programs received a 4% increase, to $7.61 billion, reflecting bipartisan support for its next fleet of planetary missions. Smaller science agencies also did well, including nearly 8% more for in-house research at the National Institute of Standards and Technology (NIST) and 6% more for the U.S. Geological Survey.

The Department of Defense's basic research program benefited from the military's gains. Funding for those programs will rise by 4%, to $2.7 billion, reversing Biden's plan to cut them by 13%.

Republicans are claiming victory for adjusting what they called a misguided imbalance in spending priorities, given rising tensions with China and Russia's invasion of Ukraine. "I insisted on dollar-for-dollar parity between defense and nondefense increases," said Alabama Senator Richard Shelby, the top Republican on the Senate appropriations panel, adding that the final version cuts "billions of dollars in wasteful proposed spending by the president."

Most Democrats agreed to the reshuffling, although Senator Patrick Leahy (D–VT), who chairs the spending panel, disputes Shelby's view that Biden's original request was excessive. Still, Leahy says the final bill "makes bold investments in critical areas that went underfunded or even neglected in the previous administration, including education, child care, health care, the environment, science and research."

The 2741-page spending package also features the return of more than 4000 earmarks—projects proposed by individual lawmakers that agencies have not requested. Academic institutions typically seek earmarks for new buildings and equipment. But this year Shelby, a master earmarker, also won $50 million for the University of Alabama (UA) to create "a permanent endowment to support the recruitment and retention of exceptional faculty in science and engineering." The money will enable UA to offer more competitive startup packages to top scientists, says a senior administrator who requested anonymity. Senator Roy Blunt (R–MO) delivered $3 million each to the University of Missouri and Missouri State University for a similar endowment targeting health and life sciences researchers, some working at a facility that already bears his name.

Earmarks are a sensitive issue for the U.S. research community: Critics say they violate the principle of using merit to select projects, whereas supporters say they level the playing field and allow institutions to address pressing needs. Congress banned them in 2010 but revived them in 2021 under new rules designed to curb excesses. Now, Shelby and Blunt have broadened what earmarks can fund. "I've never heard of using earmarks for a school's endowment," says one higher education lobbyist.

A spokesperson for UA defended its earmark, saying the endowment "supports real-world issues," and thanked Shelby for "recognizing the value" of the research being done on its campus.

Next month Biden is expected to submit his 2023 budget request to Congress. Science advocates hope it will ask lawmakers to match the aspirational spending levels for several research agencies, including NSF, DOE science, and NIST, contained in separate pending legislation that addresses China's growing economic and military might. Last week Biden urged Congress to pass the bill, which has spent nearly 2 years wending its way through both bodies.

The time frame for that bill and for appropriations is uncertain. Science lobbyists don't expect to see a final 2023 spending bill until after the November elections. If Republicans win control of one or both chambers of Congress, their disagreements with Biden's priorities could complicate a final deal. ■

A new $1 billion biomedical funding agency is born

By **Jocelyn Kaiser**

President Joe Biden has gotten his wish for a new agency to fund high-risk, cutting-edge biomedical research. Congress last week approved a 2022 omnibus spending bill that creates the Advanced Research Projects Agency for Health (ARPA-H) with a $1 billion startup investment. That's a fraction of the $6.5 billion Biden had proposed, but advocates say it's plenty to launch ARPA-H.

The bill does not resolve, however, a debate over whether to make ARPA-H a standalone agency within the Department of Health and Human Services (HHS) or part of the National Institutes of Health (NIH). Instead, it gives HHS Secretary Xavier Becerra 30 days to decide.

Biden proposed ARPA-H in 2021 as a biomedical version of the military's Defense Advanced Research Projects Agency (DARPA), famed for its nimbleness and for backing innovations like the internet. Like DARPA, it would hire program managers on short contracts who would have great freedom to solicit research ideas and swiftly fund them with milestone-driven contracts.

Acting White House science adviser Francis Collins, who stepped down as NIH director in December 2021, favors placing ARPA-H within NIH. That would allow it to draw on NIH's "brain trust," he says, as well as the practical expertise needed to quickly stand up a new agency.

But many groups have argued that ARPA-H needs to be independent to break from NIH's risk-averse funding culture and attract innovative leaders. "If it's not independent to the outside world," says Liz Feld, president of the Suzanne Wright Foundation, a pancreatic cancer research advocacy group, "all the people we need to make this effective will see it as more of the same."

Congress is split. A Senate bill introduced last week would place the agency within NIH but require it be located far from NIH's Bethesda, Maryland campus. A bill sponsored by Representative Anna Eshoo (D–CA) would make ARPA-H independent. Given that "difference of opinion," Collins says it "makes sense" to let Becerra decide.

ARPA-H will have 3 years to spend its initial $1 billion. "The expectation is that this is the beginning of a ramp" to higher funding, Collins says. For now, he says, it needs to find an interim director who can work out hiring, contracting processes, and office space. ARPA-H likely won't make awards until it has a permanent director. But Collins says that hire could come in just "a couple of months" because the position does not need Senate confirmation, and names are already "kicking around."

The ideal candidate will have broad experience in academia, industry, and philanthropy, and be familiar with translating basic discoveries into treatments, ARPA-H watchers say. Harvard University chemical biologist David Walt, a former chair of DARPA's advisory council, stresses the need for "somebody who is really a broad thinker and is not pigeonholed in their own area." The person must also be willing to divest investments that could pose a conflict. ■

ANIMAL RESEARCH

Animal care committee sues own university

Fearing harassment, members aim to block new animal rights tactic

By **David Grimm**

In an unprecedented move, members of a confidential group that oversees animal research at the University of Washington (UW) have sued their own school to block the release of their names to an animal rights organization. People for the Ethical Treatment of Animals (PETA) has been trying to obtain this information for more than a year, charging that the makeup of the university's Institutional Animal Care and Use Committee (IACUC) violates federal law. But the committee's members—citing an uptick in animal rights activism at the school, including protests at the homes of individual scientists—say they fear PETA and other animal rights organizations will use their names to target them.

"Animal rights groups have created a climate of fear at the university," says the school's IACUC chair, Jane Sullivan, who spearheaded the lawsuit. "I'm a huge fan of openness and transparency, but not when it threatens the safety of the members of my committee." She and others fear PETA's move is the beginning of a nationwide effort: The advocacy group also wants to name IACUC members at the University of Massachusetts (UMass) Amherst.

Kathy Guillermo, a senior vice president at PETA, says her organization just wants UW's committee to comply with the law. "The IACUC is the last line of defense for animals in laboratories," she says. But PETA suspects the university's committee is so biased toward research interests that it's not fulfilling its federal mandate. "The IACUC members' supposed fear of releasing their names would appear to be more about hiding a flawed process than anything else."

Every U.S. institution that receives federal money for animal research must have an IACUC with five or more members, including scientists, veterinarians, and at least one nonscientist and one person unaffiliated with the institution. That makeup is supposed to ensure that animals are properly cared for and only necessary experiments take place, according to the U.S. National Institute of Health's Office of Laboratory Animal Welfare (OLAW), which oversees these committees. Nonscientists can include ethicists and clergy members.

Most large institutions keep members' names confidential (the chair and lead veterinarian are often exceptions). That allows UW to hide the fact that its IACUC is not properly constituted, argues Lisa Jones-Engel, a senior science adviser with PETA. In September 2020, she began to file a series of public information requests with the school, asking for the names of current and former members.

Jones-Engel, a biological anthropologist at the university for 17 years, was in a unique position to challenge the IACUC: She had served on it from 2017 to 2019. Toward the end of her tenure, she filed complaints with the university alleging that the committee's nearly 20 members did not contain an ethicist and that it was "stacked" against individuals who might question animal experiments; she argued that some members considered to be nonscientists in fact had close ties to animal research. Such concerns eventually led her to leave UW and join PETA. The membership rosters she seeks will document those problems, she says.

Early last month, UW announced it would release the names, saying it was compelled by the state's open records act. Sullivan hired a lawyer and, along with four anonymous members of the IACUC, sued the school to stop the release. On 24 February, a federal judge issued a temporary restraining order, ruling that the IACUC members' fear of harassment "sharply" outweighed any "incremental knowledge" PETA was seeking.

"I can't think of any reason why the public would need to know the name of an IACUC member."

Susan Silk, formerly of NIH's Office of Laboratory Animal Welfare

Susan Silk, a former director at OLAW, agrees. She notes there are both internal and external checks on an IACUC's composition; anyone with concerns can file a complaint with either OLAW or the United States Department of Agriculture. "I can't think of any reason why the public would need to know the name of an IACUC member."

PETA's campaigns at the University of Washington include on-campus protests of monkey research, as well as demonstrations at individuals' homes and an effort to identify members of a lab animal oversight committee.

Sullivan adds that both government and private laboratory oversight bodies have found no problems with the makeup of UW's IACUC. She thinks PETA's real goal is to target committee members.

In addition to staging on-campus demonstrations at UW—whose animal use program is among the largest in the country—PETA supporters have recently protested outside the homes of two officials affiliated with the school's primate facility. Individuals wearing monkey masks held signs showing animals in cages and asking, "Do neighbors know you torture monkeys?" Sullivan says activists have left threatening emails and voice messages for university scientists and have compared IACUC members to Nazis during the committee's online meetings. "There is no question that the effect is to instill fear and terror," she says.

Jim Newman, the director of strategic communications at Americans for Medical Progress, which promotes the need for animals in labs, agrees. "You don't go to someone's house to make a general point," Newman says. "You're saying, 'We know where you live.'"

Guillermo counters that PETA's demonstrations have been "peaceful" and "legal." She says her organization "does not and has never encouraged its supporters to send or leave anything other than polite messages."

Sullivan acknowledges that neither she nor the IACUC veterinarian, whose names are both public, have been harassed at home. But she says her concern isn't only PETA—it's what more radical activists might do in response to publicity, as happened in "Pizzagate" and related incidents during the 2016 U.S. presidential election. "My biggest fear is that someone else will take matters into their own hands."

Next month, the judge who issued the restraining order may either lift the injunction or make it permanent, although PETA could then appeal. A verdict in a higher court could set a legal precedent for the state or even nationwide.

In the meantime, PETA has requested the membership of UMass Amherst's IACUC. The organization claims the committee may be improperly constituted and that its confidentiality violates the state's open meetings law.

Michael Malone, a vice chancellor at UMass Amherst, worries this strategy may spread. "If people start showing up at your house, where is that going to leave us in terms of staffing IACUCs?" he says. "People on these committees should be asking, 'Is it a good animal model?' Not, 'Am I going to be accosted on the way to the grocery store?'" ■

PHOTO: PETA

AVONET collaborator Marcus Chua measures a Salvadori's nightjar (*Caprimulgus pulchellus*).

PHOTO: MARCUS CHUA

ECOLOGY

Catalog of bird shapes yields ecological 'gold mine'

Measurements offer a new window into ecosystem health

By **Erik Stokstad**

In 2012, evolutionary biologist Catherine Sheard started an ambitious Ph.D. project: measuring the shape of every kind of passerine, or perching bird, in the world. "I thought, 'This is about 6000 species, that almost seems doable,'" Sheard says. It was, and her project catalyzed an international effort to measure all the world's birds.

Now, a team of 115 researchers from 30 countries, led by Sheard's Ph.D. adviser, Imperial College London ecologist Joseph Tobias, has published anatomical measurements of all 11,009 living bird species—not just passerines such as robins, but everything from ducks and penguins to vultures and ostriches. "It's a gold mine," says geneticist Nancy Chen of the University of Rochester, who was not involved in the project.

The open-source data set, called AVONET, debuts this month in a special issue of *Ecology Letters* along with papers describing its value for studying bird evolution and ecology, as well as the impact of changes in climate and habitat on vulnerable species. "For the first time, we are gaining a global, quantitative perspective on bird biodiversity, which is really amazing," says ecologist Brian Enquist of the University of Arizona.

Tobias drew inspiration from a massive database of plant measurements called TRY, which contains millions of records on leaf shape, chemical composition, average blooming dates, and more. By correlating these records with other types of data such as remote sensing, plant ecologists have studied a wide array of issues, including how steeply plant diversity declines when habitats are fragmented. Yet TRY has details for fewer than half of the world's 391,000 plant species, limiting its ability to answer some questions.

Assembling a complete data set for birds began to seem feasible after Sheard completed her effort, carefully wielding calipers on sometimes-fragile specimens to measure about 80 birds per day at five major museums in the United Kingdom and United States.

All told, authors contributed data from 78 collections and some field studies. On average, they measured eight to nine individuals for each species. To fill in the last few hundred missing species, Tobias networked and cold-called researchers all over the world. "By that stage it was a labor of love," he says.

The AVONET data set contains 11 morphological traits, such as beak shape and wing length, for 90,020 individual birds from 181 countries. "It's phenomenal what they've done," says Çağan Şekercioğlu, an ornithologist and conservation ecologist at the University of Utah, who created a database of bird ecological traits, including diet and habitat.

Earlier, incomplete versions of AVONET have already yielded insights. Sheard reported in 2020 in *Nature Communications* that species' geographical distributions, documented by earlier studies, correlate with flight ability, as revealed by the ratio of hand to wing length. Compared with migratory birds in temperate regions, sedentary birds in the tropics have stubbier wings, poorer flight, and more restricted ranges. That link between wing anatomy and flying range could help researchers gauge species' vulnerability to harm from habitat destruction or climate change, as poorer fliers might not be able to disperse from inhospitable environments, Şekercioğlu says.

Papers in the special issue report new findings. One shows that the evolution of flight reduced birds' reliance on weapons, such as bony spurs, likely because these defenses add extra weight. Another confirms that communities of bird species with more diversity of shapes, such as beaks specialized for niche diets, tend to have lower risks of extinction.

Other teams can apply the data to new questions. "This is really democratizing the data housed in museums," says Sahas Barve, a postdoc at the Smithsonian National Museum of Natural History. "Not only is it available for students everywhere, but it's available to scientists in the countries from where these specimens were originally taken."

Future studies can combine body shape measurements with genetic data, geographical distributions, and environmental conditions to test theories about birds' evolution and their role in ecosystems, says AVONET co-author Carsten Rahbek of the University of Copenhagen. For example, AVONET measurements can help estimate the maximum size fruit a species can eat and roughly how far it might travel before defecating the seeds—clues to which plants it might spread, and how efficiently.

Researchers could then use the data set to predict the ecological consequences of global changes, such as deforestation and warming. "This is the door to the future," Rahbek says. For example, using data on species with beaks specialized for unusual flowers, researchers could predict which plant species are at higher risk of extinction if their avian pollinators vanish. In some tropical countries, large fruit-eating birds are hunted intensely, and their loss could reduce seed dispersal. Around the world, conservation decisions "are going to have to come fast and furious," Enquist says. "Data sets like this are enabling us to anticipate what will happen and helping inform what to do."

Tobias and others plan to continue improving the data set by filling in missing data for roughly 100 species. They will also measure more individuals and add other kinds of information about life history and behaviors. For now, the data set exists as a spreadsheet in a supplemental file to a paper. Creating a community-driven database and website like TRY would require new funding, as well as mechanisms to validate newly uploaded data, such as measurements taken when researchers or volunteers capture and band living birds. "If you put it all together," Tobias says, "you could get an amazing resource." ■

PHOTO: CORNELL WATSON

THE PREGNANCY GAP

With evidence on drug safety in pregnancy sorely lacking, pressure is mounting to change the status quo

By **Jennifer Couzin-Frankel**

Kinue Beugre with her husband, Shariff McQueen, when she was 7 months pregnant.

Kinue Beugre learned she was pregnant one evening last July, 6 months after being diagnosed with lupus. Beugre, who is 22 years old, had already cycled through excruciating symptoms from the autoimmune disease: swelling that prevented her from flexing her fingers, itchiness that caused her whole body to shake, and joint pain that made climbing stairs impossible. Her husband brought meals to the second-floor bedroom of their townhouse in Durham, North Carolina, where Beugre wrote poetry and pushed ahead on a master's degree in health policy.

By the time the pregnancy test came back positive, Beugre was feeling better. A daily cocktail of drugs had tamed her body's out-of-control inflammation. The couple was thrilled about their baby-to-be and optimistic that pregnancy would proceed smoothly. "I wasn't so much worried about lupus," she says.

Weeks later, Beugre's disease flared and she faced an agonizing truth confronted by millions of people with chronic illness before her: Getting pregnant had catapulted her into a realm of unknowns. That's because for at least 90% of prescription medications available worldwide, there's little or no information on fetal safety because of a scarcity of human data.

The disparity comes in part because researchers, drug companies, and regulators largely exclude pregnant people from clinical trials, and there's little incentive to investigate how babies fare after exposure to approved medications. When it comes to medical research, there's "a sort of no-regrets policy toward pregnant people," says Ruth Faden, a bioethicist at Johns Hopkins University, "as in, 'Let's not do anything, so we won't have any regrets.'"

The problem, Faden and others say, is that without data, physicians may hesitate to recommend medications and pregnant people are often afraid to take them, sometimes with grave consequences. About 20% of pregnant people have a chronic illness, and a mother-to-be "may not survive her cancer, she may not have her diabetes well controlled, she may experience more severe postpartum depression" if her conditions go undertreated during pregnancy, Faden says. The baby can suffer, too. Flaring autoimmune diseases, for example, may trigger premature labor. On the flip side, some medicines can be harmful in pregnancy, raising the risk of congenital birth defects, developmental and other delays, or immune system problems in the baby.

Without robust evidence, "we really end up with our best-guess treatment," says Megan Clowse, a rheumatologist at Duke University and one of Beugre's doctors. "That's not good enough for our patients."

In the past few years, however, there's been growing recognition that the status quo is unacceptable and—somehow—has to change. Regulators are urging companies to learn more about their medications in pregnancy, and doctors such as Clowse are probing drug safety for their pregnant patients in any way they can. Registries track children exposed to medications before birth, looking for concerns that may crop up later. New computer models simulate blood flow, placental biology, and fetal circulation to trace the virtual passage of chemical compounds during pregnancy. More direct approaches—adding pregnant people to clinical trials and monitoring those who get pregnant while enrolled—have been slower to get off the ground.

For people like Faden, today's goal is modest: Narrow the evidence gap in drug safety between people who aren't pregnant and those who are. Even that remains a challenge.

PREGNANCY'S COMPLEX BIOLOGY can make a drug's impact hard to forecast. Physiological changes include a doubling of blood volume and increases in body fat and water. The placenta, through which nutrient-rich blood reaches the fetus and waste funnels out, changes to modulate the passage of substances across it. And an exposure's timing matters. A compound that raises the risk of limb malformations, for example, is especially hazardous in early pregnancy, when arms and legs develop.

But the intricacies of biology are not the only reason for the dearth of data. Researchers, regulators, and providers want to protect a fetus, sometimes at all costs. Patients struggle with that tension as well. "If you tell a woman, 'You're putting your child at risk,' of course they're going to stop their medication. And everyone tells them that, from their mother-in-law to their nail person," says Uma Mahadevan, a gastroenterologist at the University of California, San Francisco. Her own work has shown that inflammation from disease can be especially hazardous in pregnancy, and pregnant patients with Crohn's disease are generally at greater risk when they discontinue therapy rather than stay on it.

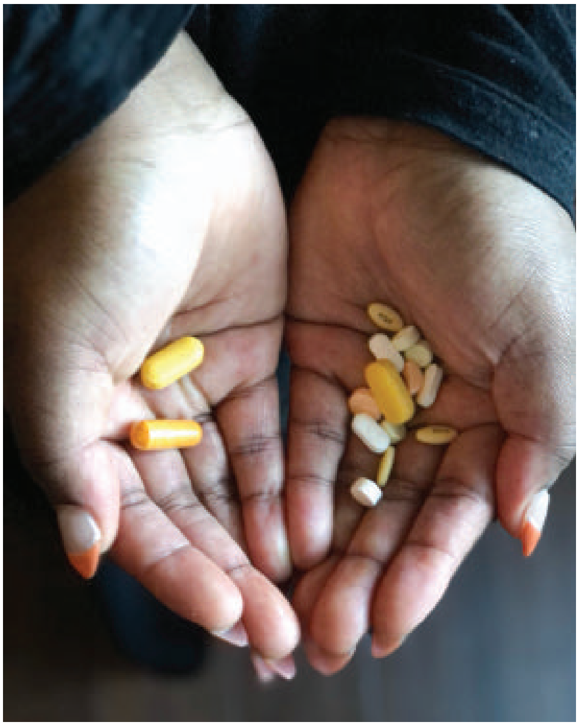

Kinue Beugre holds pills she takes to control her lupus.

PHOTO: SHARIFF MCQUEEN

History casts a dark shadow. "We don't want another thalidomide," Faden says, referring to the drug developed by a West German company in the 1950s. Thalidomide became popular worldwide for many conditions, including nausea in pregnancy, and reports soon surfaced of babies with missing or malformed limbs born to mothers who had taken it while pregnant. Thousands of babies were affected, many of whom died at or soon after birth. In 1961, thalidomide was pulled off the market. But its stain endured and left research on medications in pregnancy halting at best.

Regulatory hurdles also get in the way. The default has been to keep pregnant people out of many clinical trials for their own protection unless sponsors can justify including them. Drug trials that do include pregnant patients tend to focus on pregnancy-specific problems, such as preventing transmission of HIV to the fetus. When a pregnant person does enroll in a study that only the fetus stands to benefit from, U.S. regulations require obtaining informed consent from both parents. "There's a paternalistic approach," says Maged Costantine, a maternal fetal medicine specialist at Ohio State University, Columbus. "The old attitude has been to protect mothers from research."

At the Food and Drug Administration (FDA), the information gap came into stark relief after a 2015 rule for prescription drug labeling to address pregnancy and lactation. As revisions of dozens of labels began, regulators discovered there was little to say. "It was a data desert," says Lynne Yao, director of the division of pediatrics and maternal health. "It really highlighted the idea that we've got nothing."

Cynthia Gyamfi-Bannerman, a maternal fetal medicine specialist at the University of California, San Diego (UCSD), has seen the damage inflicted by missing information. One of the most challenging chapters of her career came last year, at a time of medical triumph: Vaccines for COVID-19 were remarkably effective, but following standard practice, drug companies had excluded pregnant people from clinical trials. And participants who happened to get pregnant after a first vaccine dose did not receive a second. Although the companies pledged to track the outcomes of several dozen participants who became pregnant, that information wouldn't be available for months.

As a result, Gyamfi-Bannerman says, she and other doctors who were part of a national task force crafting recommendations for pregnant people struggled without scientific data to guide them. Amid the uncertainty, she and her colleagues were besieged by patients afraid the vaccines could harm their fetus. There was limited evidence to reassure them.

The result of those fears: Only about 40% of pregnant people in the United States were fully vaccinated by the beginning of January, despite their heightened risk from COVID-19. Several studies that tracked the vaccines in pregnancy ultimately showed they were safe, and 8 months after authorization, the Centers for Disease Control and Prevention formally recommended that pregnant people get the shots. By then, dozens of unvaccinated pregnant people had died. (CDC, like many researchers, now favors gender-neutral language for those who are pregnant.)

A lack of data on drugs in pregnancy also poses the opposite risk: that pregnant people will take a drug that proves harmful. New Zealander Denise Astill has epilepsy, and her twin daughters, Natasha and Jazmyn, suffered serious neurodevelopmental harms from sodium valproate, a medicine Astill took while pregnant with no idea of the potential lifelong effects on her children. The twins, now 20, have autistic traits and suffer from congenital malformations and difficulty with walking and fine motor skills, as well as other medical problems. Decades passed before many countries urged against prescribing the medicine in pregnancy. And still, Astill says, little has changed. In New Zealand, "we have not got any postmarketing surveillance around pregnancy." She founded an advocacy group in 2015 to lobby for more research on drug effects and for support for families like hers.

"I honestly don't understand," says Rebecca Bromley, a pediatric neuropsychologist at the University of Manchester who has assessed and cared for affected children like Natasha and Jazmyn. "As soon as a medication is approved, someone [should be] appointed to immediately follow the first births." This should include the launch of multiple independent studies, she says, to provide comprehensive information.

OVER THE YEARS, some researchers have tried to fill in the data gaps, one puzzle piece at a time. At UCSD, perinatal epidemiologist Christina Chambers was dissatisfied with the usual evidence on drugs in pregnancy, which generally comes from animal research. Companies often give pregnant animals doses many times higher than those prescribed to people, looking for effects on fetal growth and sur-

vival, and major malformations in the offspring. The result, Chambers says, is better than nothing, but such findings often don't translate neatly to people.

So in the late 1990s, she and her colleagues launched a nationwide registry that has enrolled thousands of pregnant people who use various medications (or sometimes no medication at all) to treat asthma, high blood pressure, depression, and many other conditions. Participants agree to be called regularly and to release their and their children's medical records, allowing researchers to hunt for differences between those who are and are not on a certain medicine. Among other findings, the MotherToBaby registry reported that flu vaccines and adalimumab, a drug used to treat rheumatoid arthritis and Crohn's disease, aren't associated with poor pregnancy outcomes.

Some physicians also have established registries for patients with specific conditions. Mahadevan has so far recruited 2000 pregnant people for PIANO, the Pregnancy Inflammatory bowel disease And Neonatal Outcomes registry, and among other findings has documented that when disease flares, patients have a higher risk of miscarriage. Her registry recently added annual questionnaires to gather information on participants' children until they turn 18.

Registries, Chambers says, provide an early warning that drugs may cause problems—or reassurance that their risks appear low. But registries can also suffer from sluggish enrollment of patients on a hodgepodge of therapies and limited funding. "It takes 20 years to collect vaguely useful data" from registries, says Clowse, who despite her misgivings runs one for rheumatic disease in pregnancy that includes about 800 people. She sees an urgent need to give better guidance, especially for the sickest patients navigating pregnancy—which in her practice includes those with lupus who are at high risk of very early delivery.

One was Beugre, whose health took a dive when she was 17 weeks pregnant. An immunosuppressant called azathioprine, which is used in pregnancy, risked harming her liver, so Clowse cut the dose. Soon after, at her sister's October birthday party, Beugre's skin began to itch again. Within days her eyes were swollen each morning when she awoke, and one day she couldn't open her right eye at all.

Beugre raced to the hospital, where she received high-dose steroids. And then came the difficult question: What next? Inflammation from the disease, she was told, could spread to the placenta and kill her fetus. Clowse pinned her hopes on another lupus treatment: rituximab, a monoclonal antibody first approved in 1997 that tamps down the immune system. But Clowse was nervous, too, because rituximab can interfere with the development of an important class of immune cells, called B cells, in the second half of pregnancy. Beugre was teetering on that line.

"We had to make a pretty rushed decision," Clowse says. "Every week of delay would put the baby at risk." Beugre received two infusions, which helped, and Clowse fervently hoped they would carry her patient through the rest of her pregnancy. After she'd delivered her baby boy, Beugre could get another course of rituximab—or newer medications, whose effects on a fetus are unknown.

For Clowse, walking that tightrope, as she'd done many times before, was immensely frustrating. Her registry can't offer quick facts about new drugs. "We need some pregnancy-focused drug trials to answer important questions on new therapeutics," she says.

AS COSTANTINE HAS LEARNED firsthand, however, running traditional trials in pregnant patients is littered with obstacles. He focuses on preeclampsia, a life-threatening condition of pregnancy marked by soaring blood pressure that's cured only by delivering the baby, often far too early. Desperate to find a better treatment, Costantine began more than a decade ago to study a cholesterol drug called pravastatin, first approved in the United States in 1991. It looked promising: Experiments in rats by several groups showed that pravastatin reduced the risk of preeclampsia without apparent downsides. Beginning in 2012, FDA permitted pilot studies, and over time Costantine enrolled 40 pregnant patients who'd had severe preeclampsia in a prior pregnancy, leaving them at high risk of a repeat. Nine of 20 in the placebo group developed the condition, compared with two of 20 in the treatment group. But the numbers were too small to declare the drug effective.

In 2018, Costantine received funding from the National Institutes of Health for a definitive trial of more than 1500 patients, but FDA told him he could include no more than 50. The agency's concern, Costantine says, stemmed from decades-old data from rats suggesting that at doses six to 10 times higher than what he is studying, the drug can affect part of the developing brain. FDA officials have asked Costantine to repeat the rat studies, which he says would cost more

ILLUSTRATION: ALEX SG ADAPTED BY C. AYCOCK/*SCIENCE*

than $1.5 million and add little value given that much can be learned from the babies born in his studies.

Those children are now 2 to 7 years old and by all appearances are healthy; Costantine presented an update on 30 of them at a meeting in February, covering verbal, spatial, motor, and other skills. In fact, he says, the children born to patients who got pravastatin have had slightly better outcomes, which he speculates are partly because they were less likely to be born prematurely. Costantine feels caught in a catch-22 that, he says, encapsulates much of the field. “The best way to treat the mothers is to have good evidence,” he argues. “The only way to get this evidence is to do research.”

FDA’s Yao declined to speak about the pravastatin case but acknowledged the agency’s balancing act. “We just don’t want to do the wrong thing,” she says, adding that solid preclinical data are crucial before testing a therapy on pregnant people. At the same time, “I fully understand the flip side: Being too conservative can also hurt.”

Moving pravastatin, a drug approved long ago, into pregnancy trials has been hard enough. With experimental therapies, the hurdles loom even higher. There can be “a huge liability issue” if a pregnant study volunteer has a baby with a birth defect, for example, says Marie Teil, a clinical trialist with experience in research ethics. She was hired by UCB Biopharma, a pharmaceutical company in Belgium, to help it learn how to study its products in pregnant and breastfeeding people. And there are ethical concerns, too. Teil questions whether experimental treatments, whose effectiveness is by definition unproven, should be offered in pregnancy given unknown risks to the fetus.

But, Teil says, other avenues are available for learning. After joining the company in 2013, she focused on a UCB therapy called certolizumab pegol (Cimzia), first approved in 2008 for rheumatoid arthritis and other immunologic conditions. Lab findings suggested it might not cross the placenta. To prove that in patients and assess whether the therapy could pass into breast milk, the company launched two human studies of pregnant and breastfeeding patients, enrolling only those already taking certolizumab. Researchers tested participants’ blood during pregnancy, umbilical cord blood at birth, breast milk, and blood from the baby, by heel prick, at 4 and 8 weeks of age.

The studies, which each included about 16 volunteers, took more than 2 years. Ultimately, researchers found that “no to minimal” drug reached breast milk or crossed the placenta. Europe and the United States added this new evidence to the drug label. Now high on Teil’s to-do list is keeping track of those who become pregnant during any UCB drug trial, though she says she does not expect the company would continue them on experimental medications. Those patients, who may have been exposed to a new therapy in early pregnancy, could offer valuable clues about drug safety for pregnant people and their children.

UCB’S EVOLVING POLICIES come amid a broad effort to spark change. The push began in earnest about 5 years ago, with the congressionally mandated Task Force on Research Specific to Pregnant Women and Lactating Women (PRGLAC, pronounced “preg-lac”). PRGLAC’s members, generally physicians and scientists with expertise in maternal and pediatric health, many at government agencies, delivered a laundry list of recommendations to the Secretary of Health and Human Services and Congress in September 2018. The list included no longer requiring paternal consent for participation in studies, funding programs to study drugs that pregnant and breastfeeding people commonly use, developing a strategy to mitigate liability risk for companies, requiring investigators to justify excluding pregnant and breastfeeding patients from clinical trials, and boosting lab research on drugs in pregnancy.

“We really end up with our best-guess treatment. That’s not good enough for our patients.”

Megan Clowse, Duke University

Some of PRGLAC’s bullet points call for a sea change. But others are what advocates describe as “low-hanging fruit,” and some are beginning to be implemented. In 2021, moved partly by PRGLAC’s urgings, the National Institute of Child Health and Human Development committed $27.5 million over 5 years to create the Maternal and Pediatric Precision in Therapeutics (MPRINT) Hub. Among the projects it supports is one mining electronic health records to study opioid exposure in pregnancy and babies after birth, and another that will probe antibiotics in mouse models and pregnant and breastfeeding patients and their babies. MPRINT and the Bill & Melinda Gates Foundation are also working to develop “virtual pregnancy” computer models that forecast how drugs will behave.

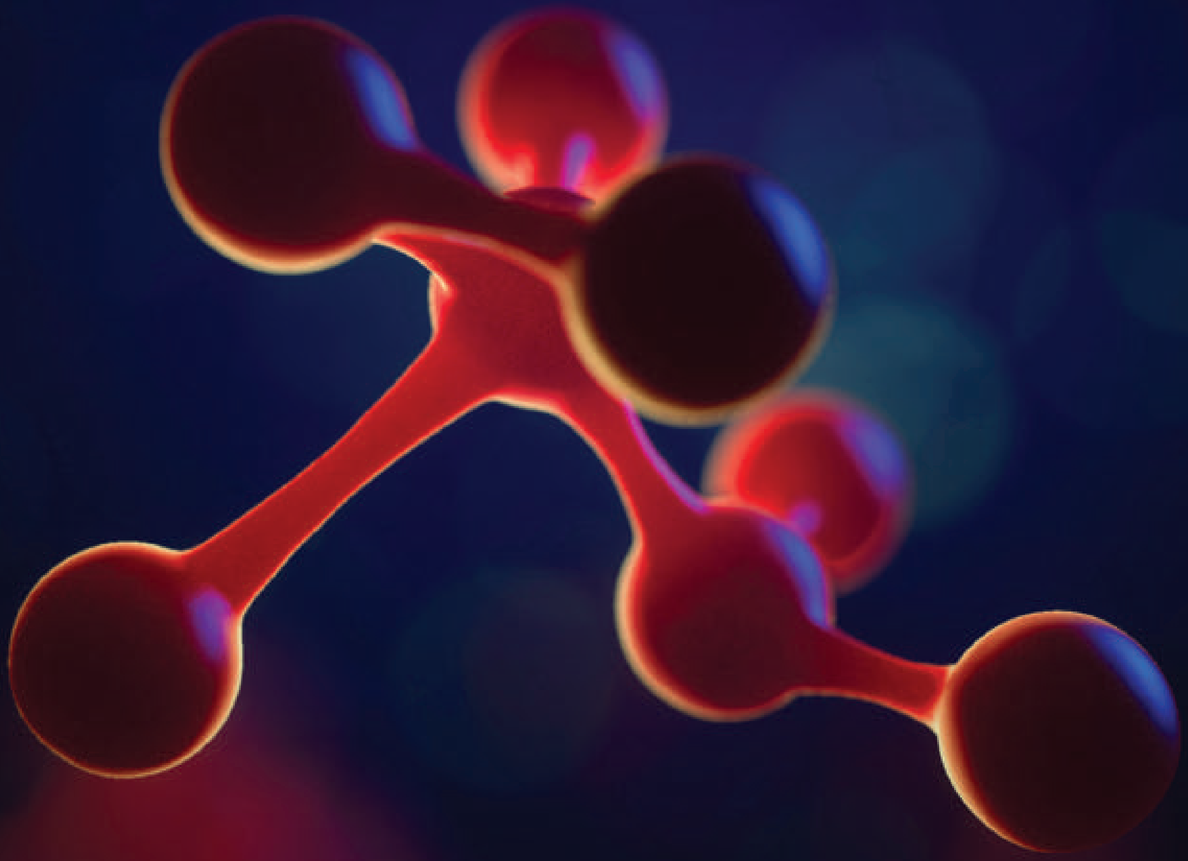
FDA, meanwhile, recommended in 2018 draft guidance that when a patient gets pregnant during a trial, she be followed and even offered the chance to continue on the study medication if that had been her assigned group. “We have tried very hard to improve in that space,” Yao says of FDA’s policies. Among other efforts, the agency is also trying to craft strategies to better communicate risk and research findings to physicians caring for these patients.

But guidance can accomplish only so much, Costantine believes, noting the exclusion of pregnant people from COVID-19 vaccine and treatment trials. “There needs to be a cultural shift in viewing pregnant people as medically complex, rather than vulnerable, needing protection,” he says.

And even if the drug safety picture sharpens in the coming years, pregnant patients will still face the intensely personal question of how much risk to accept. “What do women actually care about?” Clowse asks. It’s widely agreed that certain medications, such as thalidomide and valproate, should be avoided in pregnancy. But at times, some risk to the fetus may be worth an outsize benefit to the mother. Protecting her can help, and even save, her baby, too. Clowse says her patients “want their child to be pretty healthy. Do they have to be perfect? Are there children who are perfect?”

For Beugre, the balancing act has continued through pregnancy, but winter proved far less fraught than fall. She’s taking the steroid prednisone, whose long-term effects on a child’s health, if any, are unknown—like nearly all drugs’. Her dose, 12.5 milligrams per day, is a notch higher than ideal in pregnancy. Mindful of the tightrope, Beugre chose not to increase it when a modest flare caused hair loss and joint pain. But she’s not comfortable lowering the dose, either, aware that more serious flares can harm her fetus. “I have to go off my own intuition of what I can handle,” she says.

And she has arrived at the finish line: As *Science* went to press, Beugre had passed 38 weeks, with an induction to end her high-risk pregnancy just days away. Noble Saleem Beugre-McQueen’s arrival is imminent, and his mother is relieved that she can no longer tell whether her achy joints are due to illness or to the final days of pregnancy. ■




INSIGHTS

POLICY FORUM

CARBON MARKETS

Crediting agricultural soil carbon sequestration

Regional consistency is necessary for carbon credit integrity

By **Emily E. Oldfield**[1], **Alison J. Eagle**[1], **Rachel L. Rubin**[2], **Joseph Rudek**[1], **Jonathan Sanderman**[2], **Doria R. Gordon**[1,3]

Rapid expansion of the voluntary carbon market is evidenced by the number of carbon registries and private companies that have recently published measurement, reporting, and verification (MRV) protocols to pay farmers for sequestering carbon in agricultural soils as a climate change mitigation strategy. Our review of 12 publicly available soil organic carbon (SOC) MRV protocols from major registries and other crediting organizations reveals important differences in their approaches to measurement and estimation of SOC and net greenhouse gas (GHG) reductions and to key accounting issues, such as additionality and permanence (see the table) (*1*). These differences pose a risk of nonequivalent credit creation, which would undermine confidence in the integrity of crediting programs. We outline ideas for a regionally consistent crediting framework that could help establish consistent, transparent, and robust criteria for generating SOC credits and addressing risks and challenges associated with the voluntary SOC market.

The protocols reviewed focus on credits generated from atmospheric drawdown of carbon dioxide (CO_2) with sequestration in the soil through management changes in arable systems, as opposed to credits for avoided or reduced emissions such as avoided land conversion or nitrogen management. One carbon credit represents one tonne of CO_2 removed from the atmosphere or the equivalent amount of a different GHG (CO_2e). These credits could be used for a variety of purposes, such as fulfillment of corporate sustainability goals. Most crediting programs have launched within the past 1 to 3 years. Some have issued and sold credits, whereas others will verify and sell the first round of credits within the next year. However, it will likely be 5 years from now (when the first round of soil sampling after project initiation occurs) before we can empirically evaluate whether and how much carbon was actually sequestered.

In addition to an expanding voluntary carbon market landscape, the European Union (EU) and the United States are considering policy mechanisms to scale up the adoption of emissions-reducing and carbon-storing agricultural practices. For example, in 2021 the US Senate passed the Growing Climate Solutions Act, which is intended to help farmers benefit from a market-based system that incentivizes climate-friendly agricultural practices, and the EU published technical guidance on how to increase removals of CO_2 from the atmosphere through carbon farming (*2*). Agreements reached on Article 6 at the 2021 UN Climate Change Conference set rules for the international sale of credits that could spur investment in land-based climate-mitigation projects.

PHOTO: USDA NRCS SOUTH DAKOTA

---

[1] Environmental Defense Fund, Washington, DC, USA. [2] Woodwell Climate Research Center, Falmouth, MA, USA. [3] Department of Biology, University of Florida, Gainesville, FL, USA. Email: eoldfield@edf.org

Young corn grows out of crop residue from the prior season. Crop residue retention and no-tillage can improve soil health and lead to soil carbon accrual.

Setting criteria to define a high-quality SOC credit can help ensure that lower standards are not rewarded within the marketplace, to the benefit of farmers, registries, project developers, and credit buyers.

MEASUREMENT CHALLENGES

Quantifying and verifying real, net changes in GHGs associated with soils is difficult, primarily because of challenges of measuring changes in SOC. Management-induced changes in SOC over annual to decadal time frames may not be detectable with sufficient confidence given the slow rate of accrual and the large spatial variation of SOC. Also, programs that issue credits based on increased SOC stocks or avoided CO_2 emissions must account for potential increases in other potent GHGs—especially nitrous oxide (N_2O) and methane (CH_4)—that might accompany changes in cropland management. Current crediting programs differ in how and whether they address net GHGs (see the table).

Approaches used in existing protocols to measure carbon stock changes include direct sampling, indirect measurement (such as remote sensing), model estimates (by using process-based or empirical models), and hybrid approaches that combine direct sampling and modeling (see the table). Protocols that involve soil sampling generally require sampling to 30 cm (with some recommending sampling down to 1 m). Including only the top 30 cm horizon fails to account for a substantial portion of the total SOC stock and may not adequately detect vertical SOC redistribution that results from certain practices (*3*). Protocols also vary in baseline definition, ranging from static baselines established through initial sampling to dynamic baselines modeled over time to incorporate fluctuations due to past management decisions and current weather. These methodological differences can yield widely differing net sequestration rates estimated for a given practice over the lifetime of a project. There is currently no standard by which to judge the accuracy of the methods used.

Differences in scale at which these measurement approaches are deployed has substantial implications for both project feasibility (necessary time and financial resources) and accuracy of quantified GHG reductions. Project scale (field, farm, or aggregated set of fields) is rarely defined in existing protocols; when specified, minimum requirements range from roughly 4 to 400 ha. Direct measurement of SOC is time intensive and expensive, which limits the number of samples that can be feasibly collected. For example, the direct measurement approach under Australia's Carbon Farming Initiative relies on field-scale soil sampling, the high cost of which (roughly \$22/ha) presents a sizable barrier to project implementation of this smaller project-based approach (their goal is to reduce costs to <\$3/ha) (*4*). Furthermore, an approach based only on soil sampling generates credits over longer time frames (generally 3 to 5 years), resulting in delayed payments to farmers.

Protocols that use model-based estimates of net-GHG reductions that result from shifts in management practices avoid the need for intensive sampling and can issue credits annually. Process-based biogeochemical models can, in theory, be deployed at different scales from subfield to farm to region. However, limited precision associated with model inputs can increase uncertainty at the site level; thus, process-based models generally do not provide accurate estimates for a single field, especially without detailed site-specific data (*5*). Uncertainty is inversely related to scale in process-model estimates of SOC changes, with uncertainties of ~20% at a US national scale growing to 600 to 700% at the site scale (*6*). Thus, models alone are inadequate for soil C estimation at small scales unless there has been considerable calibration in the areas and for the crops over which the model is used (*5*).

ACCOUNTING INCONSISTENCIES

In addition to technical challenges of SOC quantification, credits must account for additionality (emission reductions exceed those that would have happened without the crediting system), leakage (emission-reducing practices under the program do not cause increased unaccounted emissions elsewhere), reversals (protection against subsequent losses due to changing practices or unforeseen climate impacts such as fire, flood, or drought), and permanence (maintenance of sequestered carbon stocks over a specified time frame, often defined as 100 years). These types of accounting issues arise in many environmental policy domains and can risk undermining the desired climate benefits.

Published protocols address these issues, but with varying thresholds (see the table). For example, in addressing additionality, some protocols use a performance standard approach that requires projects to demonstrate that practices affecting SOC storage are not already being implemented on a percentage of land area containing a project: >50% in the case of the Soil Enrichment Protocol (SEP) by the Climate Action Reserve (CAR), >20% for Verra's VM0042, and >5% for Gold Standard's SOC Framework (GS-SOC). Thus, a project might be eligible for credits under CAR SEP but not under GS-SOC. Further, the Nori and Regen Network protocols allow for "look-back" periods, meaning that farmers can earn credits for practices adopted up to 10 years before the current year. BCarbon offers credits for any carbon sequestered after the onset of sampling with no performance requirement. Less stringent thresholds for additionality, including look-back periods, are often presented as ways to include early adopters in crediting programs, an important consideration for fairness; however, the climate benefit of the credit is diluted.

A REGIONAL APPROACH

A regional accounting and verification framework can help address shortcomings of smaller projects, enabling systemic policy and institutional changes at scale (*7*). The approach we outline is an adaptation of the jurisdictional approach that is gaining traction for generating credits for tropical forest protection that quantifies emissions reductions relative to a baseline for an entire economic sector across a jurisdiction (*7*). Most SOC MRV protocols allow for aggregation of field- or farm-based projects; however, how to ensure that fields are appropriately aggregated is rarely defined and is at the discretion of project developers and registries.

Under a regionally consistent framework, the regional unit could be a biophysically defined agroecological zone that has similar soils, climate, and agricultural potential or constraints (for example, the US Department of Agriculture's Major Land Resource Areas; similar land classifications exist in the EU, India, China, and elsewhere). This approach relies on administration by a public or private entity (a government agency or an independent nonprofit organization) that would set criteria for the generation and independent verification of SOC credits to which protocols and project developers must adhere. Standardization across regions would provide consistency, and yet incorporate parameters specific to each region as needed. Traditional modes of carbon credit generation would still operate in the voluntary market within each region (with registries responsible for protocols and with farmers and project developers receiving and selling issued credits), but to meet regional standards, registries and project developers must follow regional criteria. In this way, the framework we propose is akin to certification.

This regional framework will likely require substantial public investment, but there are opportunities for public-private partnerships as well. For example, a regional approach effectively aligns with corporate sustainability commitments to source and supply commodities that demonstrate improved sustainability because typical supply chains of food

How protocols address critical structural considerations

Details are available in the supplementary materials.

CATEGORY	ALBERT CC	AUS-DV	AUS-SM	BCARBON	CAR SEP	FAO GSOC	GS-SOC	NORI
APPLICABLE GEOGRAPHY	Canada	Australia	Australia	US	US	International	International	US
MEASUREMENT APPROACH	Model	Model	Sampling	Hybrid	Hybrid	Hybrid	Hybrid	Model
PERMANENCE REQUIREMENT (YEARS)	20	25 (discounted) or 100	25 (discounted) or 100	10	100 or tonne-year accounting	8	20	10
ACCEPTABLE LEVEL OF UNCERTAINTY (%)	Reflected in model	Reflected in model	Probability of exceedance > 60%	10	15	Depends	20	Depends
ADDITIONALITY ASSURANCE								
Performance Standard (implementation defined percentage of land area)					✓		✓	
Legal requirement (specific practices not required by law)					✓		✓	
Proof of additionality using CDM Additionality tool							✓	
Initial modeling projections show SOC sequestration above BAU						✓		✓
Project activities must be new and additional to BAU	✓	✓	✓					
MANAGING FOR REVERSALS								
Percentage of credits go to buffer pool	✓	✓	✓	✓	✓	✓	✓	
Restricted credits due to higher reversal risk								✓
ACCOUNTING FOR LEAKAGE	✓	✓	✓	✓	✓	✓	✓	✓
INCLUSION OF OTHER GHGs IN OVERALL ESTIMATES	✓	✓	✓		✓	✓	✓	

Alberta CC, Alberta Quantification Protocol for Conservation Cropping; AUS-DV, Australian Carbon Credits (Carbon Farming Initiative-Estimating Sequestration of Carbon Using Default Values) Methodology Determination; AUS-SM, Australian Carbon Credits (Carbon Farming Initiative-Measurement of Soil Carbon Sequestration in Agricultural Systems) Methodology Determination; BCarbon, Protocol for Measurement, Monitoring, and Quantification of the Accrual of Below-Ground Carbon Over Time; CAR SEP, Climate Action Reserve Soil Enrichment Protocol; FAO GSOC, Food and Agriculture Organization GSOC MRV Protocol; GS-SOC, Gold Standard Soil Organic Carbon Framework Methodology; Nori Croplands Methodology; Regen Network Methodology for GHG and Co-Benefits in Grazing Systems; VM17, Verra Adoption of Sustainable Land Management; VM21, Verra Soil Carbon Quantification Methodology; VM42, Verra Methodology for Improved Agricultural Land; CDM, Clean Development Mechanism; BAU, business as usual; GHGs, greenhouse gases.

companies are not traceable to the level of a specific farm. A more transparent and effective regional accounting system would add credibility to investment in land-based mitigation strategies.

Measuring change at the regional scale

Monitoring change in GHGs across a region has several advantages. The greater accuracy of models at regional scales resolves uncertainty generated at the field or project level (*6*). Similarly, direct measurements of soil are also more efficient at larger scales as the variance per unit area decreases at broader spatial scales, thus enabling greater efficiency for soil sampling designs (*8*). Using an ensemble of process-based models as a component of SOC MRV at large scales could increase confidence in estimates of mean changes in net SOC versus accounting for changes in soil carbon on a project-by-project basis (*9*). The generalized emissions factors commonly used across all published protocols to estimate other potent GHGs (such as N_2O and CH_4) are coarse estimates that are much more applicable and accurate at a regional scale than when applied at smaller project scales (*10*). Last, emerging remote sensing approaches show promise in tracking rates of soil erosion at broad scales, which can help constrain estimates of soil carbon sequestration (*11*).

The monitoring and measurement approach under a regional framework would involve credit estimation by using models but with a network of sites (both within and outside of participating farms) to allow for verification of BAU and improved management practice scenarios specific to stratified zones within each region. This could account for variation in climate and soil that influence crop productivity and soil carbon sequestration potential (such as soil type, aridity index, annual growing degree days, and annual temperatures). The regional agent would help coordinate soil sampling at project inception and beyond to recalibrate models. The advantage of sampling counterfactual sites is that it allows for validation and verification of BAU practices, whereas current protocols rely solely on modeled estimates of SOC dynamics under past practices. Countries such as the United States and in Europe have soil monitoring programs that could be adapted to suit this purpose. Intensive measurements at a network of sites will help to advance our mechanistic understanding of the processes involved in SOC turnover and storage, which will help improve model parameterization and projections of SOC stocks into the future (*12*, *13*). In addition, a larger set of auxiliary sites located within stratified regions could provide a reduced suite of measurements (crop yields, SOC, bulk density, and N balance) to allow for true out-of-sample model validation. This would help to facilitate research and data collection with the goal of creating high-quality, open-access datasets for model calibration and benchmarking that use standardized guidelines for SOC, N_2O, and CH_4 sampling and measurement (*14*).

Accounting to ensure benefits

The agroecological zones could be the regional frame in which the regional agent

REGEN NETWORK	VM17	VM21	VM42
International	International	International	International
Remote sensing	Model	Sampling	Hybrid
25	30	30	30
20	15	10	15
			✓
			✓
	✓	✓	
✓			
✓	✓	✓	✓
✓	✓	✓	✓
✓	✓	✓	✓

maintains an inventory of land use, production, agricultural practices, and carbon project locations, providing data that would allow for more transparent and consistent accounting of additionality and leakage. In addition to increased consistency, this approach would transfer data needs (such as baseline, additionality, and leakage calculations and prevention of double-counting) from registries and project developers to the regional agents, which could potentially save on transaction costs and allow greater revenue to be passed to producers. This framework could address structural inconsistencies among current SOC MRV protocols. Calibrated remote-sensing products for tracking cover cropping and tillage practices within a region could be used to monitor against perverse incentives (tilling a no-till field to participate in the market) and provide more data on rates of practice adoption.

Accounting at a larger scale and collating relevant and publicly available statistics (such as land use, crop yields, and climate variables) to administer each region would also achieve appropriate scale to track annual land use change and crop failure, which are critical for assessing leakage and climate-induced reversals in carbon storage. For example, regions that are more prone to drought may require larger buffer contributions to account for potential losses of the carbon stock. Analyzing annual and long-term yield averages across the entire region (including participating and nonparticipating cropland) would allow determination of whether there is systematic loss in yields (sustained declines >5% of regional average) resulting from the crediting program.

Most protocols account for leakage within a "leakage area," which lacks consistency across protocols and runs the risk of insufficient spatial accounting to ensure against increased emissions outside of the project boundary. Accounting for changes in crop yields and land use at a regional level will help ensure against leakage or emission shifts within the region (*15*). Communication among regions will be essential to account for possible production shifts. Addressing leakage beyond the region (such as deforestation in the tropics) requires continued vigilance and monitoring, with solutions that need to reach a broader scale.

A regional approach could potentially help implementation of improved practices at scale and also benefit a greater diversity of farms by enabling access to and participation in markets. Coordination within regional networks of farms and farmers can provide education and technical support, with barriers to market entry addressed through coordinated financial assistance based on farm size to reduce the risks of practice adoption (subsidizing costs of equipment or seed); or setting a price floor on carbon. These efforts could help build a sufficiently large carbon stock at the regional scale that is buffered by events that might occur at the individual field level (for example, a farmer opting out or tilling their field). The longer a program persists, the more the risk of reversals or loss of the SOC stock from weather or management can be pooled across farmers and time periods (*7*).

GROWING MOMENTUM

Given the current heterogeneity in approaches represented by the voluntary carbon market, this regional approach will help to provide consistency and transparency to ensure that credits are equivalent and high quality. Nesting a regional approach under standards set at the US level (or other aggregated geographies) would ensure greater consistency at a jurisdictional scale should the current voluntary market be subsumed into a compliance-based approach in the future. This approach is not without challenges and limitations. Setting criteria and establishing regional oversight will necessitate substantial public and private investments (such as managing infrastructure required for regional monitoring networks, administrative needs to support data collection and collation, and coordinating outreach and assistance programs). Engaging stakeholders in the process of setting regional criteria is critical as is representing the diverse array of regional growing operations, regional supply chains, and those organizations that will implement the criteria (*15*). Coordination across and within these multiple levels can be a difficult and lengthy process. Despite the challenges, this framework effectively aligns with corporate supply chain accounting and potentially with national GHG inventory efforts.

Given the growing momentum behind agricultural soil carbon sequestration, a unified framework will instill greater confidence in generated credits that help achieve climate goals. ■

REFERENCES AND NOTES

1. E. E. Oldfield *et al*., "Agricultural soil carbon credits: Making sense of protocols for carbon sequestration and net greenhouse gas removals" (Environmental Defense Fund, 2021); https://www.edf.org/sites/default/files/content/agricultural-soil-carbon-credits-protocol-synthesis.pdf.
2. European Commission, Directorate-General for Climate Action, G. Radley *et al*., *Setting Up and Implementing Result-Based Carbon Farming Mechanisms in the EU: Technical Guidance Handbook* (Publications Office of the European Union, LU, 2021); https://data.europa.eu/doi/10.2834/056153.
3. D. A. Angers, N. S. Eriksen-Hamel, *Soil Sci. Soc. Am. J*. **72**, 1370 (2008).
4. Australian Government Department of Industry, Science, Energy and Resources, "Technology investment roadmap: First low emissions technology statement—2020" (Australian Government Department of Industry, Science, Energy and Resources, 2020), p. 44.
5. C. Tonitto, P. B. Woodbury, E. L. McLellan, *Environ. Sci. Policy* **87**, 64 (2018).
6. S. M. Ogle *et al*., *Glob. Change Biol*. **16**, 810 (2010).
7. S. Schwartzman *et al*., *Environ. Res. Lett*. **16**, 091001 (2021).
8. R. T. Conant, K. Paustian, *Environ. Pollut*. **116** (suppl. 1), S127 (2002).
9. C. Riggers *et al*., *Geoderma* **345**, 17 (2019).
10. S. J. Del Grosso, H. T. Gollany, M. Reyes-Fox, in *Synthesis and Modeling of Greenhouse Gas Emissions and Carbon Storage in Agricultural and Forest Systems to Guide Mitigation and Adaptation* (John Wiley & Sons, 2016), pp. 89–110.
11. E. A. Thaler, I. J. Larsen, Q. Yu, *Proc. Natl. Acad. Sci. U.S.A*. **118**, e1922375118 (2021).
12. M. A. Bradford *et al*., *Biogeochemistry* **156**, 19 (2021).
13. S. M. Ogle, K. Butterbach-Bahl, L. Cardenas, U. Skiba, C. Scheer, *Curr. Opin. Environ. Sustain*. **47**, 28 (2020).
14. K. Paustian *et al*., *Carbon Manag*. **10**, 567 (2019).
15. M. von Essen, E. F. Lambin, *Front. Ecol. Environ*. **19**, 159 (2021).

ACKNOWLEDGMENTS

This research was supported through gifts to the Environmental Defense Fund from the High Meadows Foundation for postdoctoral fellowships; the Bezos Earth Fund; and Arcadia, a charitable fund of Lisbet Rausing and Peter Baldwin. We thank S. Hamburg, S. Kerr, J. Lavallee, and R. Lubowski for feedback on this manuscript. We also thank representatives from each crediting organization for reviewing our summary of their protocol.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl7991

10.1126/science.abl7991

EVOLUTION

The rapid tempo of adaptation

Selection in fruit flies leads to fast adaption to seasonal changes

PHOTO: SATOSHI KURIBAYASHI/MINDEN PICTURES

By **Ary H. Hoffmann**[1] *and* **Thomas Flatt**[2]

Understanding the tempo and mode of adaptive change is a long-standing problem in evolutionary biology (*1*). Although evolutionary dynamics are traditionally thought to be much slower than ecological changes, growing evidence shows that adaptation can occur at a pace comparable to that of ecological changes, as seen in Darwin's finches and other species (*2*). Fast adaptability is particularly important when organisms are facing rapidly changing environments. Populations can cope with these conditions either by evolving a "generalist" or an environmentally "plastic" phenotype, or by genetically and phenotypically adapting to each environmental change in what is known as "adaptive tracking" (*3*). Yet, classical theory predicts that adaptive tracking may be rare and unlikely to maintain genetic variation (*4–6*). On page 1246 of this issue, Rudman *et al.* (*7*) provide compelling evidence for seasonal adaptive tracking in field populations of *Drosophila melanogaster*—the fruit fly.

The fruit fly is a laboratory workhorse for examining the genetic basis of core biological processes. However, this unassuming insect has also played an important role in many ecological and evolutionary studies, including those on genetic variation in natural populations (*8, 9*). This work and classical studies in ecological genetics, such as those on snails and moths, show how genetically determined phenotypes like wing morphs can mediate adaptation to environmental changes such as climate and predation (*10*). Rudman *et al.* build on findings in *Drosophila pseudoobscura* from the 1940s by geneticist Theodosius Dobzhansky, who observed seasonal fluctuations in the frequencies of chromosomal inversions in natural populations due to temporally varying selection (*11*). Contrary to classical theory, these and other studies suggest that selection is often sufficiently strong and variable across time to maintain different genetic variants (*6*), yet detailed empirical demonstrations of adaptive tracking are rare.

The study by Rudman *et al.* begins to fill this major gap between theory and observation and extends earlier work documenting changes in specific genetic markers and heritable ecological traits in fruit fly populations. The authors followed genomic changes in 10 independent replicate field cages of flies (100,000 per cage) exposed to changing weather across the season. In parallel, they documented monthly changes in quantitative traits after rearing flies in a "common garden" laboratory setting that is standardized to exclude environmentally induced and parental effects that would cause trait changes. The surveyed traits represent major components of fitness, including developmental rate and egg production, and three stress resistance traits: survival under cold, starvation, and desiccation. All of the observed fitness-related traits evolved in parallel among flies in all the cages, implying that selection and not random genetic change was responsible for trait evolution. However, the rate and direction of the phe-

---

[1] School of BioSciences, Bio21 Institute, University of Melbourne, Melbourne, Australia. [2] Department of Biology, University of Fribourg, Fribourg, Switzerland.
Email: ary@unimelb.edu.au; thomas.flatt@unifr.ch

Research shows that fruit flies (*Drosophila melanogaster*) can adapt very rapidly to seasonally changing environments, in a phenomenon called "adaptive tracking."

notypic evolution varied depending on the trait. For example, chill coma recovery, an indicator of cold resistance, increased steadily toward winter, whereas desiccation resistance increased, then plateaued, and finally decreased. Notably, the rates of phenotypic evolution were extremely fast as compared to those seen in other animal species in the wild (*12*). Such a rapid evolutionary tempo is consistent with adaptive tracking, which requires maximal rates of evolution (*3*).

By testing for parallel changes in 10 replicate cages (each containing thousands of individuals), performing DNA sequencing for groups of 100 flies with a sequence coverage of >100-fold, and sampling every few weeks, Rudman *et al.* could detect temporal and parallel changes in the frequencies of gene variants (i.e., alleles) among replicates—a clear telltale of selection. As expected under strong fluctuating selection, many parallel frequency shifts changed direction over time. The genomic architecture underlying these changes indicates that many large blocks of the genome responded to selection, because of pervasive nonrandom associations between sites under selection with those unrelated to the selection response. This might reflect the use of experimental populations with only four generations of genetic recombination (i.e., the process that reshuffles gene variants) among the fly strains that were used to establish the populations. Thus, because recombination of DNA had a limited opportunity to break down these associations, the study cannot resolve fine-grained changes at small genomic regions. Despite this, the marked parallelism of the response to fluctuating selection represents strong evidence for seasonal adaptation through adaptive tracking. Further work is required to identify underlying causative changes.

An interesting aspect of the Rudman *et al.* study is the lack of evidence for inversions—when a segment of a chromosome is reversed—to be driving genetic changes. This seems to run counter to evidence for the importance of such rearrangements in rapid seasonal adaptation, as originally observed by Dobzhansky (*11*) and further supported by more recent studies (*4*). However, the frequencies of common inversions are quite low in natural *Drosophila* populations in Pennsylvania, where the flies used by Rudman *et al.* were from. This may help explain the lack of inversion contributions in their study. An open question is how the tempo and mode of seasonal adaptation depend on the populations of origin and their genetic makeup, including the prevalence of inversions.

Rudman *et al.* establish that fly populations can rapidly track seasonal changes. This might be expected on theoretical grounds as *Drosophila* populations are often very large with much genetic variation, and their evolution is not limited by rare mutations. Under such conditions, even weak selection can be effectively strong and drive a fast evolution. However, classical theories predict that adaptive tracking can be hindered by factors such as a reduction of fitness because of the time lag in adaptation after each environmental change, and other factors such as segregation load (fit heterozygous genotypes producing unfit offspring) and pleiotropy (genetic changes simultaneously affecting multiple traits that reduce fitness) (*3, 5*). Why these expectations are not borne out in real-world *Drosophila* populations remains unclear.

A recent comparison of genomic and quantitative genetic data suggests that balancing selection, possibly involving temporally varying selection, can contribute more to maintaining genetic variation than previously thought (*13*). In support of the underappreciated importance of fluctuating selection, previous genomic analyses have documented pervasive seasonal allele frequency changes in several fruit fly populations (*4*, *14*), and seasonal heritable changes have also been documented in flies for cold resistance (*8*). Emerging theories suggest that the conditions for fluctuating selection to maintain variability might be less stringent than previously assumed (*5*). The work by Rudman *et al.* lends strong credence to the feasibility of this mechanism and highlights the core issue (*15*) of how genetic variability can be maintained. ■

REFERENCES AND NOTES

1. G. G. Simpson, *Tempo and Mode in Evolution* (Columbia Univ. Press, 1944).
2. N. G. Hairston Jr. *et al.*, *Ecol. Lett.* **8**, 1114 (2005).
3. A. M. Simons, *Proc. Biol. Sci.* **278**, 1601 (2011).
4. H. E. Machado *et al.*, *eLife* **10**, e67577 (2021).
5. M. J. Wittmann, A. O. Bergland, M. W. Feldman, P. S. Schmidt, D. A. Petrov, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E9932 (2017).
6. A. M. Dean, C. Lehman, X. Yi, *Genetics* **205**, 1271 (2017).
7. S. M. Rudman *et al.*, *Science* **375**, 1246 (2022).
8. A. A. Hoffmann, A. R. Weeks, *Genetica* **129**, 133 (2007).
9. J. R. Adrion, M. W. Hahn, B. S. Cooper, *Trends Genet.* **31**, 434 (2015).
10. E. B. Ford, *Ecological Genetics* (Chapman and Hall, 1975).
11. T. Dobzhansky, *Genetics* **28**, 162 (1943).
12. A. P. Hendry, T. J. Farrugia, M. T. Kinnison, *Mol. Ecol.* **17**, 20 (2008).
13. B. Charlesworth, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 1662 (2015).
14. A. O. Bergland, E. L. Behrman, K. R. O'Brien, P. S. Schmidt, D. A. Petrov, *PLOS Genet.* **10**, e1004775 (2014).
15. R. C. Lewontin, *The Genetic Basis of Evolutionary Change* (Columbia Univ. Press, 1974).

10.1126/science.abo1817

ASTRONOMY

Unifying repeating fast radio bursts

Mysterious high-energy radio bursts are found to share certain characteristics

By **Manisha Caleb**

Indiscernible to the human eye, emanating from distant galaxies, are immensely energetic flashes of radio waves called fast radio bursts (FRBs) (*1*). These radio signals are both fast and explosive, releasing as much energy in a millisecond as the Sun releases over several days. Their transient nature, mysterious origins, and the sheer amounts of energy they exude on such short time scales make them fascinating. Most FRBs are one-off events, but a small subset has been observed to repeat (*2*). The diversity in the observed properties of FRBs has raised the possibility of differences even within repeating FRBs, especially displayed in their varying frequency and polarization characteristics (see the figure). On page 1266 of this issue, Feng *et al.* (*3*) present a unified characterization of repeating FRBs based on the frequency evolution of the FRB polarization.

Polarization is a fundamental property of electromagnetic waves and is typically used to probe the presence of magnetic fields. The rotation measure (RM) parameter quantifies the rotation of the electromagnetic waves by the foreground material they move through, whereas the polarization position angle provides information about the emission mechanism. Overall, polarization measurements of FRBs provide information about their origin as well as the immediate environment of the source producing them (*4*). Repeating FRBs offer astronomers the opportunity to study the polarization properties of FRBs in exquisite detail (*5*) with different telescopes and over various frequency ranges. Of all the known FRBs, few have had their polarization measured by astronomers. This is primarily because the data products are too large to be effectively stored and processed as a result of current computational limitations.

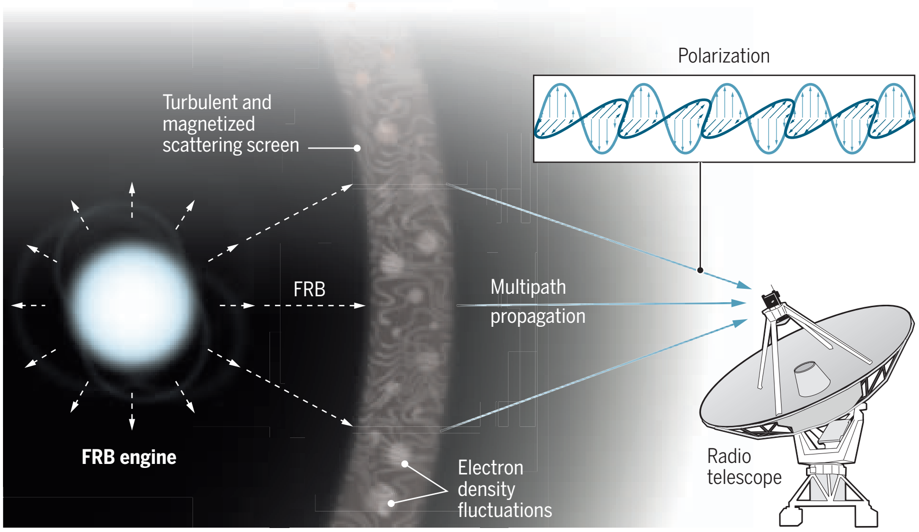
Sydney Institute for Astronomy, School of Physics, The University of Sydney, NSW 2006, Australia.
Email: manisha.caleb@sydney.edu.au

Additionally, not all telescopes have polarization acquisition capabilities.

The very first polarization measurement of a repeating FRB was made for FRB 20121102A, which showed a 100% linear polarization with an extremely high RM of 10^5 rad/m^2 in the 4- to 8-GHz range and no discernable linear polarization in the 1.0- to 2.4-GHz range (*4*). Pinpointing the precise location within the puny dwarf host galaxy showed the origin of the bursts of electromagnetic waves to be colocated with a bright, persistent radio wave counterpart. This possible association hints at a compact object in a supernova remnant or a pulsar wind nebula as a possible source of the FRB (*6*). Notably, the RM of this FRB has decreased in its subsequent bursts between 2017 and 2019, which suggests a changing magnetized environment possibly related to the diffuse and expanding persistent source (*7*).

The mysterious repeating fast radio bursts

Radiation from repeating fast radio burst (FRB) sources appears to rotate as they pass through magnetically active areas in the universe. Variation in this rotation is a result of the different ray paths taken by each burst. A larger variation can mean the FRB source is relatively young and resides in a more turbulent magnetized environment.



The discovery of a repeating FRB prompted fervent discussions on the ubiquity of repeaters in the Universe and even the possibility of all FRBs being repeaters. The localization of a second repeating source, FRB 20180916B, to the outskirts of a massive spiral galaxy presented a stark contrast to the puny dwarf galaxy hosting the first repeater FRB 20121102A (*8*). Compared with FRB 20121102A, FRB 20180916B had a much smaller RM with a magnitude of 114 rad/m^2 and was not seen to be associated with persistent radio emission. This discovery hinted at a possible disparity even among repeaters (*8*).

This hypothesis was supported by an even more recent observation of FRB 20200120E, which is localized to a globular cluster in the Messier 81 galactic system. The absence of a persistent radio counterpart as bright as the one from FRB 20121102A and an even smaller average RM of 37 rad/m^2 for the FRB support the notion of different environments and/or ages among repeating FRBs (*9*). A match to FRB 20121102A also was discovered in the repeater FRB 20190520B. Both FRBs reside in dwarf galaxies and are associated with persistent radio emission sources (*10*). Additionally, similar to FRB 20121102A, no linear polarization was detected in the 1.0- to 1.5-GHz range for FRB 20190520B, but an average RM of 2759 rad/m^2 was measured in the 4- to 8-GHz range (*10*). Altogether, this suggested a possible evolutionary link between the source properties and its surrounding plasma.

To study this possible link, Feng *et al.* analyzed the polarization properties of a sample of five prolific repeating FRBs observed with the Five-hundred-meter Aperture Spherical radio Telescope (FAST) and the Green Bank Telescope (GBT) to understand their origins and environments. Their examinations of the degrees of polarization versus observing frequency band for each repeat burst across all five FRBs reveal a correlation such that repeat FRBs are seen to have a lower degree of polarization when observed at a lower frequency band. The authors attribute this correlation to the distribution of the measured RMs of FRBs. The distribution, or scatter, is the extent to which the RMs of the repeat bursts from each of the five FRBs are likely to vary about the average value for that FRB. The scatter in the RM is postulated to be caused by multipath transmission in an inhomogeneous magneto-ionic environment, leading to a decrease in the polarized intensity with frequency in a process known as depolarization.

The bursts from the five repeaters are consistent with being nearly 100% linearly polarized and becoming depolarized through transmission processes, which can be characterized by the RM scatter parameter. A larger RM scatter is attributed to a younger, more active source with a surrounding medium that is more complex in terms of turbulence, density, and/or magnetization compared with other FRBs. These objects with a more scattered RM have larger RM magnitudes and are more likely to be associated with persistent emission. For example, the larger scatter in RMs attributed to FRB 20121102A and FRB 20190520B and the small scatter in RM ascribed to FRB 20180916B could be related to the presence and absence of persistent radio sources, respectively. These persistent counterparts may be directly related to the object producing the FRB and could provide vital clues to its nature.

If the scatter in RM is indeed caused by the multipath propagation effect in individual FRBs, one would also expect this effect to manifest as a temporal broadening of the pulse profile. Notably, the observed temporal broadening in individual FRBs is found to positively correlate with the corresponding scatter in RM. This positive correlation is consistent with the hypothesis that both the RM scatter and temporal broadening arise from the same plasma screen—likely a supernova remnant or pulsar wind nebula surrounding the FRB engine. Feng *et al.* provide a distinct way of probing the complexity of the FRB environment and possibly the evolutionary stage of a repeating source. FRBs from persistent radio sources may be discovered by searches of superluminous supernovae and long gamma-ray bursts. Future observations will test whether these FRBs with larger RM scatters and compact persistent radio counterparts are younger than other FRBs. The field of FRBs is still evolving and promises decades of exciting astrophysics to follow. ■

REFERENCES AND NOTES

1. D. Thornton *et al.*, *Science* **341**, 53 (2013).
2. M. Caleb, E. Keane, *Universe* **7**, 453 (2021).
3. Y. Feng *et al.*, *Science* **375**, 1266 (2022).
4. D. Michilli *et al.*, *Nature* **553**, 182 (2018).
5. K. Nimmo *et al.*, *Nat. Astron.* **5**, 594 (2021).
6. S. Chatterjee *et al.*, *Nature* **541**, 58 (2017).
7. G. H. Hilmarsson *et al.*, *Astrophys. J. Lett.* **908**, L10 (2021).
8. B. Marcote *et al.*, *Nature* **577**, 190 (2020).
9. F. Kirsten *et al.*, *Nature* **602**, 585 (2022).
10. C.-H. Niu *et al.*, https://arxiv.org/abs/2110.07418 (2022).

10.1126/science.abo2353

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

METABOLISM

Fueling the fire of adipose thermogenesis

Specialized fat tissue generates heat and holds the potential to counter metabolic diseases

By **Christian Wolfrum**[1]
and **Zachary Gerhart-Hines**[2,3,4]

Adipose thermogenesis is the energy lost as heat through metabolism of macronutrients in specialized fat cells, or adipocytes (*1*). This program is activated by neuronal, hormonal, and metabolic cues in response to cold temperature, dietary excess, and time of day. The capacity of thermogenic adipocytes to dissipate macronutrient calories as heat is vital to body temperature defense in rodents and, likely, human infants. In adult humans, this tissue serves as a catabolic "sink" to burn off excess blood macronutrients, and its activity is linked to improved body weight regulation, glucose and lipid control, and cardiometabolic health (*2*). Stimulating adipose thermogenesis is an appealing strategy to counteract metabolic disruptions, such as obesity and type 2 diabetes, but has yet to yield a pharmacotherapy that is both safe and efficacious. However, deeper interrogation of heat-producing macronutrient pathways and the emerging roles of macronutrients in shaping thermogenic capacity offer the possibility of finally unlocking this distinct biology.

Adipose tissue is composed of three different types of adipocytes distributed heterogeneously in depots throughout the body. White adipocytes store calories in the form of triglycerides (TGs) in a single large lipid droplet. This anabolic function as an energy reservoir also serves to protect against whole-body lipotoxicity by sequestering potentially harmful lipid species from the circulation. By contrast, brown and beige (or brite) adipocytes are rich in mitochondria and have multiple, smaller lipid droplets. These are thermogenic adipocytes owing to their capacity to consume glucose, lipids, and other substrates to fuel catabolic processes that generate heat. Similar to white adipocytes, thermogenic adipocytes protect against lipotoxicity and macronutrient excess but do so through oxidation of substrates instead of storage. Underscoring the role for thermogenic adipose tissue as a catabolic sink, consumption of the tricarboxylic acid (TCA) cycle intermediate, succinate, by brown and beige adipocytes protected mice from succinate-induced pathological inflammation of the liver (*3*).

The canonical means of heat production in mouse and human adipocytes occurs through the mitochondrial transporter, uncoupling protein 1 (UCP1) (*1*). Upon cold exposure, free fatty acids (FFAs) released from intracellular lipid droplets or from TG-rich lipoproteins in the blood enter mitochondria and bind to UCP1 (*1*). UCP1 then shuttles protons from the intermembrane space into the mitochondrial matrix and thus uncouples substrate oxidation (and proton release) from driving adenosine triphosphate (ATP) synthesis (see the figure). The potential energy lost by dissipation of the proton motive force is released as heat. Uncoupling substrate oxidation from ATP synthesis facilitates increased activity through the electron transport chain (ETC), which requires reduced substrates to donate electrons so protons can be pumped back into the intermembrane space. This demand for reduced ETC substrates is met by elevated TCA cycle activity, which uses intermediates from macronutrient catabolism. Increased flow through these exergonic (energy-releasing) pathways and the ETC are the basis for thermogenesis. Ultimately, brown and beige adipocytes fuel this heightened catabolic metabolism by consuming macronutrients from the blood, thereby directly linking adipocyte thermogenesis to whole-body glucose and lipid control and weight regulation (*2*).

In addition, heat production in adipose tissue also occurs through futile metabolic circuits where substrates, such as metabolites or ions, are cycled between the same starting and ending states, culminating in a net loss of energy as heat. One well-described futile metabolic circuit involves the hydrolysis of TGs in lipid droplets to FFAs and glycerol in the cytosol followed by anabolic esterification that consumes ATP to re-form TGs (*4*). Another futile cycle involves creatine phosphorylation and dephosphorylation, which is ascribed to skeletal muscle as a means of rapidly replenishing ATP during exertion. However, in activated thermogenic adipocytes, creatine is continuously cycled between phosphorylated and dephosphorylated states, resulting in heat production through repeated ATP hydrolysis and subsequent accelerated oxidation of substrates through the ETC (*5*). Additionally, the sarcoplasmic and endoplasmic reticulum (SER) calcium ATPase 2b (SERCA2b) pump drives the repeated exchange of calcium between the SER and the cytosol (*6*). Similar to creatine cycling, calcium cycling thermogenesis is driven by continuous ATP consumption and, ultimately, increased ETC activity. Within thermogenic adipose depots there are heterogeneous subpopulations of fat cells (*7*), and it is not known in which subpopulations these various cycles occur. Nor is it known if such cycles are exclusive hallmarks of thermogenic adipocytes or if subtypes of white adipocytes can contribute as well. Given the many different futile cycles, future studies should establish their coordinated regulation, relative importance, and quantitative contribution to total adipose thermogenesis.

Coordinated transport of macronutrients and ions into and out of mitochondria plays a critical role in adipose thermogenesis, so it is important to complete a functional map of adipocyte mitochondrial metabolite carriers (the SLC25A family). This class of transporters, which includes UCP1, are the gatekeepers of metabolite, lipid, and ion exchange across the inner mitochondrial membrane. Yet, the substrate is unknown for ~30% of SLC25A family members, leaving major gaps in knowledge about biological programs such as thermogenesis (*8*). Moreover, even among the SLC25A transporters with known substrates, there is insufficient resolution of flux information in adipocytes to generate an interactive network of coordinated metabolite exchange. Underscoring this point, the transporter, SLC25A44, was recently de-orphanized as the branched chain amino acid (BCAA) carrier in brown adipocytes, providing additional mechanistic understanding of how substrates are supplied for the ETC and thermogenesis (*9*). Given the associations between circulating BCAAs and metabolic disease, the robust consumption of these macronutrients by thermogenic adipocytes provides another example of how this tissue acts as a sink to benefit metabolic health.

---

[1]Institute of Food, Nutrition, and Health, ETH Zurich, Schwerzenbach, Switzerland. [2]Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, DK. [3]ADIPOSIGN: Center for Adipocyte Signaling, University of Southern Denmark, DK. [4]Embark Biotech ApS, Copenhagen, DK.
Email: christian-wolfrum@ethz.ch; zpg@sund.ku.dk

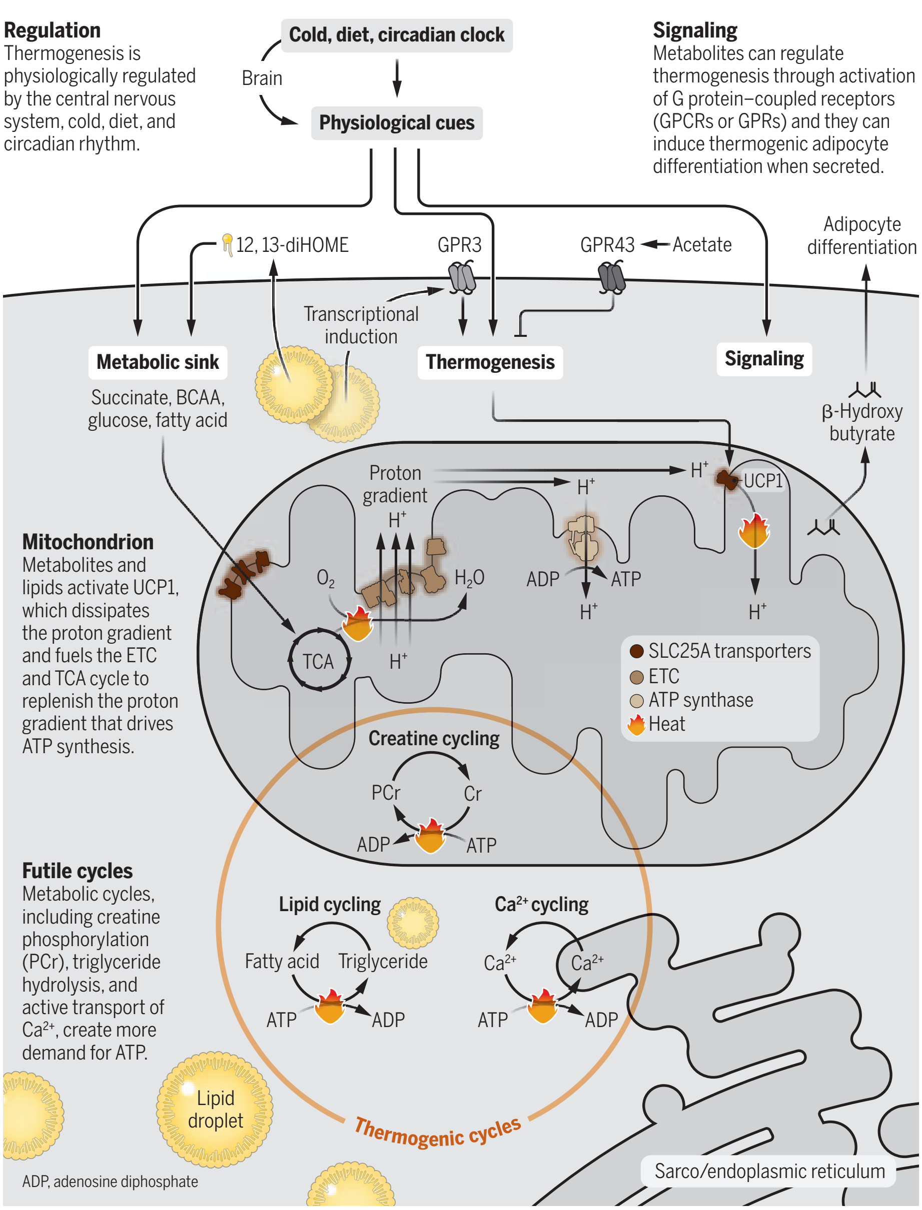
Beyond their roles as catabolic fuels for futile cycles and mitochondrial oxidation, macronutrients also control adipocyte thermogenic capacity as signaling molecules. This can be both autocrine and paracrine, between adipocytes and other cell populations within a depot. For example, a lipolytic signal from lipid droplets in brown adipocytes triggers the expression of the constitutively active G protein–coupled receptor 3 (GPR3), which cell autonomously induces ETC activity and thermogenesis (*10*). Other lipid signals are secreted, such as 12,13-diHOME, which is released from cold temperature–activated brown adipocytes and increases fatty acid uptake to directly boost the supply of oxidative substrates (*11*). Thermogenic adipocytes also use macronutrient signals to trigger proliferation. The ketone body, β-hydroxybutyrate, is released from mature beige adipocytes and drives precursor cells to differentiate into thermogenically competent adipocytes, thereby strengthening the energy-expending capacity of the tissue (*12*).

Macronutrient signals are also used as a means of tempering adipose thermogenesis. Acetate, released locally within the tissue, signals to adipocytes through the short-chain fatty acid receptor, GPR43, to inhibit thermogenic output (*7*). This negative regulation may contribute to the decreased prevalence of brown adipocyte activity in adult humans (*7*). Given the selectivity and cell surface accessibility of their receptors, targeting the signaling functions of metabolites and lipids offers an appealing, untapped avenue to therapeutically exploit adipose thermogenesis. However, this approach likely requires the optimization of agonists or antagonists engineered to have improved stability and longer half-lives in circulation than the endogenous lipid and metabolite signals, so as to elicit and maintain a meaningful degree of efficacy.

Macronutrients shape adipose thermogenesis as direct oxidative substrates, mediators of futile cycles, and signaling molecules. Yet key mechanistic questions remain regarding how metabolites and lipids control heat production and how this regulation translates to human physiology. This is due, in large part, to the experimental limitations of inferring dynamic, metabolic information from gene or protein profiles or from metabolomic assessment of known or previously annotated compounds. These approaches certainly have merit and have produced countless insights about adipose thermogenesis. However, in vivo flux metabolomics techniques now allow unbiased tracking of the flow of an isotopically labeled atom in a macronutrient of choice throughout organs of free-moving animals, including humans (*13*). This methodology offers the possibility of interrogating thermogenic pathways in adipose tissue with previously inaccessible resolution. In vivo flux analysis is already being used to explore how brown adipocytes shuttle one such thermogenic substrate, glucose, into different metabolic pathways during acute or prolonged activation by cold temperature in mice (*14*). This revealed that glucose is rapidly used to synthesize fatty acids, which are then oxidized to fuel thermogenesis (*14*). Another in vivo flux study, comparing different tissues in mice, demonstrated that brown adipocytes power lipid synthesis from glucose through the pentose phosphate pathway (*15*), a glycolytic shunt used to produce reduced nicotinamide adenine dinucleotide phosphate (NADPH) for anabolic metabolism. Future investigations will need to incorporate additional labeled macronutrients and explore contexts of genetic, pharmacological, and pathological perturbations to adipose

Metabolic control of adipose thermogenesis

Once activated, thermogenic adipocytes are metabolic sinks that consume macronutrients, such as glucose, fatty acids, succinate, and branched-chain amino acids (BCAAs), from the blood. Their heightened catabolic metabolism, through uncoupling protein 1 (UCP1) activation and futile cycling, drives demand for reduced electron transport chain (ETC) substrates from intermediates of the tricarboxylic acid (TCA) cycle. Increased flow through the TCA cycle and ETC, for adenosine triphosphate (ATP) synthesis, is the basis of thermogenesis.



Regulation
Thermogenesis is physiologically regulated by the central nervous system, cold, diet, and circadian rhythm.

Signaling
Metabolites can regulate thermogenesis through activation of G protein–coupled receptors (GPCRs or GPRs) and they can induce thermogenic adipocyte differentiation when secreted.

Mitochondrion
Metabolites and lipids activate UCP1, which dissipates the proton gradient and fuels the ETC and TCA cycle to replenish the proton gradient that drives ATP synthesis.

Futile cycles
Metabolic cycles, including creatine phosphorylation (PCr), triglyceride hydrolysis, and active transport of Ca^{2+}, create more demand for ATP.

ADP, adenosine diphosphate

GRAPHIC: V. ALTOUNIAN/SCIENCE

thermogenesis to create an integrated overview of how catabolic adipose tissue influences organismal health.

Yet perhaps the most substantial impact that in vivo flux analyses could offer is to bridge investigations of human physiology and therapeutic potential. The use of nonradioactive isotopes in such approaches is conducive to safe application in humans of various ages and levels of health. Flux analyses could identify new metabolic biomarkers that may be used in the clinic to noninvasively track adipose thermogenesis. Additionally, this technology could lead to the discovery of more energy-dissipating pathways that physiologically contribute to thermogenic activity in adipocytes or pinpoint which node in a particular catabolic pathway would be maximally efficacious to target pharmacologically for treatment strategies.

Whether the health benefits of adipose thermogenesis can be transformed into widespread therapeutic application to counteract metabolic diseases remains unclear. Increasing adipocyte energy expenditure in an older, obese, and diabetic target population poses challenges for recruiting enough thermogenic adipose tissue, which functionally decreases with age and metabolic disruption. Furthermore, thermogenic pharmacotherapies must avoid safety risks, such as excessive increases in heart rate, blood pressure, and body temperature. In addition, activating macronutrient consumption and oxidation by brown and beige adipocytes would likely require pairing with an appetite suppressant, such as glucagon-like peptide 1 receptor agonists, to prevent homeostatic food intake compensation. Ultimately, with new insights into how adipocyte thermogenesis is fueled and advanced research tools in hand, the focus should now be to translate the activity of energy-expending adipocytes to clinical fruition. ■

REFERENCES AND NOTES

1. B. Cannon, J. Nedergaard, *Physiol. Rev.* **84**, 277 (2004).
2. T. Becher *et al.*, *Nat. Med.* **27**, 58 (2021).
3. E. L. Mills *et al.*, *Nat. Metab.* **3**, 604 (2021).
4. A. J. Brownstein *et al.*, *Rev. Endocr. Metab. Disord.* **23**, 121 (2022).
5. L. Kazak *et al.*, *Cell* **163**, 643 (2015).
6. K. Ikeda *et al.*, *Nat. Med.* **23**, 1454 (2017).
7. W. Sun *et al.*, *Nature* **587**, 98 (2020).
8. J. J. Ruprecht, E. R. S. Kunji, *Trends Biochem. Sci.* **45**, 244 (2020).
9. T. Yoneshiro *et al.*, *Nature* **572**, 614 (2019).
10. O. Sveidahl Johansen *et al.*, *Cell* **184**, 3502 (2021).
11. M. D. Lynes *et al.*, *Nat. Med.* **23**, 631 (2017).
12. W. Wang *et al.*, *Cell Metab.* **30**, 174 (2019).
13. C. R. Bartman *et al.*, *Nat. Metab.* **3**, 896 (2021).
14. S. M. Jung *et al.*, *Cell Rep.* **36**, 109459 (2021).
15. Z. Zhang *et al.*, *Nat. Metab.* **3**, 1608 (2021).

10.1126/science.abl7108

CELL BIOLOGY

Copper-induced cell death

Excess copper causes mitochondrial protein aggregation and triggers a distinct form of cell death

By **Martha A. Kahlson** *and* **Scott J. Dixon**

Copper and other trace metals are essential for life. However, it is important that these metals are present at the right amount in cells. Too little metal can impair the function of important metal-binding enzymes, and too much metal can overwhelm a cell, leading to death. In humans, mutations that cause excess copper accumulation are life-threatening, and yet there may exist a window such that more focused increases of intracellular copper can be used to selectively kill cancer cells (*1*). A better understanding of how copper accumulation causes cellular toxicity is therefore of great interest. On page 1254 of this issue, Tsvetkov *et al.* (*2*) reveal that copper toxicity involves the disruption of specific mitochondrial metabolic enzymes, triggering an unusual mechanism of cell death. This mechanism may explain the pathology associated with genetic copper overload disorders and suggest new ways to harness copper toxicity to treat cancer.

> **"These new results may also invigorate studies exploring the use of copper to treat cancer."**

Stressors—including DNA damage, protein misfolding, or disruption of the cytoskeleton—can all result in cell death owing to activation of the apoptosis pathway. In the past decade, advances have been made in understanding how essential trace metals can exert toxic effects on mammalian cells. An emerging theme is that these various metals can kill cells in ways that do not require the apoptosis pathway. For example, excess zinc can trigger nonapoptotic cell death by inhibiting adenosine triphosphate (ATP) synthesis (*3*, *4*). Iron can catalyze the formation of toxic membrane lipid peroxides that are the cardinal feature of ferroptosis, a distinct form of nonapoptotic cell death (*5*). As for copper, evidence from yeast and mammalian cells links copper toxicity to altered function of mitochondria (*6*, *7*). How exactly copper disrupts mitochondrial function and how this leads to death have remained unclear.

The study by Tsvetkov *et al.* makes strides in addressing how copper accumulation disrupts mitochondrial function. They demonstrate that excess copper can selectively perturb a set of metabolic enzymes that have a distinctive posttranslational modification, lipoylation. Lipoylation involves attachment of the small sulfur-containing metabolite lipoic acid to a substrate protein. This flexible and chemically reactive appendage enables catalysis by swinging between different subunits within an enzyme complex, donating or accepting electrons. There are only four known lipoylated enzymes, all of which are found in the mitochondria, including dihydrolipoamide *S*-acetyltransferase (DLAT), a subunit of the pyruvate dehydrogenase complex (*8*). Tsvetkov *et al.* find that copper toxicity can be suppressed by genetic disruption of enzymes required for the synthesis of lipoic acid or of individual lipoylated enzymes themselves. They also reveal a previously unknown function for the mitochondrial enzyme ferredoxin 1 in lipoic acid synthesis, confirming that disruption of this enzyme inhibits copper-dependent toxicity (*7*). Lipoic acid itself is found to bind copper, and it is proposed that this causes the lipoylated mitochondrial enzymes to aggregate within this organelle in a toxic manner (see the figure).

Mechanistically, how the aggregation of copper-bound, lipoylated mitochondrial enzymes cause a nonapoptotic form of cell death remains to be fully fleshed out. Tsvetkov *et al.* show that this lethal mechanism does not involve a profound disruption of respiration, an essential energy-generating process, and that copper-induced death cannot be blocked by chemical inhibitors of lipid peroxidation or reactive oxygen species accumulation. This distinguishes copper-induced cell death from both zinc-induced cell death and ferroptosis. So how do cells die? One possibility is that by inhibiting multiple lipoylated enzymes in parallel, copper abruptly alters the concentrations of many metabolites simultaneously to induce an acute lethal shock. Alternatively, the ag-

---

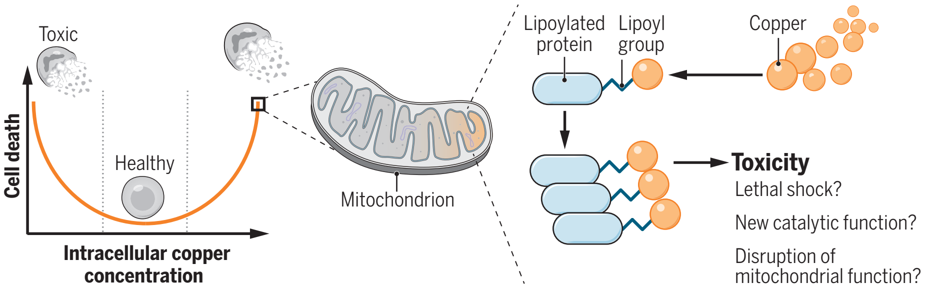
Department of Biology, Stanford University, Stanford, CA, USA. Email: sjdixon@stanford.edu

gregated, copper-bound enzymes could acquire a new catalytic activity, leading to the production of a toxic metabolite that the cell is not equipped to handle. Apart from this, the protein aggregates could kill cells by disrupting the function of other mitochondrial enzymes, such as those involved in the synthesis of iron-sulfur clusters, which are essential cofactors for other enzymes (*7*).

Still, why cell death triggered by excess copper, or other metals, does not channel the cell toward an established cell death mechanism like apoptosis is unclear. One tantalizing possibility for copper toxicity is that, in addition to promoting mitochondrial protein aggregation, this metal actively inhibits enzymes necessary for the execution of apoptosis (*9*). In this way, excess copper may clear a path toward executing cell death by an alternative mechanism.

Copper triggers toxic mitochondrial protein aggregation

Too little or too much copper is toxic to cells. When there is too much copper, such as through treatment of cells with elesclomol, copper accumulates in the mitochondria. This results in aggregation of lipoylated proteins, including dihydrolipoamide *S*-acetyltransferase (DLAT), leading to cell death.



Tsvetkov *et al.* show a strong link between copper toxicity and mitochondrial activity. Although copper itself does not greatly affect mitochondrial respiration, the toxicity of this metal is enhanced many fold in actively respiring cells. How respiration promotes cell death is not clear. Cells in a more respiratory state may express higher amounts of lipoylated enzymes, which can then create more aggregates. Another possibility is that during active respiration, higher metabolic flux through lipoylated enzymes like DLAT may increase their copper-binding affinity, leading to greater aggregation. Regardless, the link between active mitochondrial respiration and copper-induced cell death harken to the idea that copper, like iron and zinc, participates in forms of cell death best described as "cell sabotage," in which the activity of an essential metabolic pathway can drive a cell toward death when the smooth operation of that pathway is disrupted (*10*).

These findings also suggest that pharmacological inhibition of mitochondrial respiration may be a strategy, albeit one fraught with the potential for adverse side effects, to combat Wilson's disease, a hereditary copper overload disorder that results from a mutation in the copper exporter ATP7B and is lethal if untreated.

These new results may also invigorate studies exploring the use of copper to treat cancer. In their studies, Tsvetkov *et al.* use elesclomol, a synthetic molecule that can bind copper present in the environment and bring it into the cell to induce cell death. Elesclomol has been tested in human clinical trials for the treatment of epithelial cancers, albeit with limited success (*11*, *12*). The new mechanistic knowledge reported by Tsvetkov *et al.* suggests that elesclomol might work best in cancers that express high amounts of lipoylated mitochondrial proteins and that are highly respiratory. This approach may be especially useful for cancers that are naturally resistant to apoptosis, enabling a new way to kill cancer cells by exploiting the distinct action of this metal. ■

REFERENCES AND NOTES

1. E. J. Ge *et al.*, *Nat. Rev. Cancer* **22**, 102 (2022).
2. P. Tsvetkov *et al.*, *Science* **375**, 1254 (2022).
3. K. E. Dineley, T. V. Votyakova, I. J. Reynolds, *J. Neurochem.* **85**, 563 (2003).
4. W. Du *et al.*, *Cell Rep.* **37**, 109848 (2021).
5. V. E. Kagan *et al.*, *Nat. Chem. Biol.* **13**, 81 (2017).
6. R. K. Blackman *et al.*, *PLOS ONE* **7**, e29798 (2012).
7. P. Tsvetkov *et al.*, *Nat. Chem. Biol.* **15**, 681 (2019).
8. E. A. Rowland, C. K. Snowden, I. M. Cristea, *Curr. Opin. Chem. Biol.* **42**, 76 (2018).
9. S. Tardito *et al.*, *J. Am. Chem. Soc.* **133**, 6235 (2011).
10. D. R. Green, *Cell* **177**, 1094 (2019).
11. S. J. O'Day *et al.*, *J. Clin. Oncol.* **31**, 1211 (2013).
12. B. J. Monk *et al.*, *Gynecol. Oncol.* **151**, 422 (2018).

ACKNOWLEDGMENTS

S.J.D. is a co-founder of Prothegen, Inc., and a scientific advisor to Ferro Therapeutics and Hillstream Biopharma.

10.1126/science.abo3959

MOLECULAR BIOLOGY

One residue—one function

The histone H3.1 variant deposited at replication forks docks DNA repair machinery

By **Frédéric Berger**

DNA is packaged around nucleosomes formed by two copies of each core histone, H2A, H2B, H3, and H4. Essential components of chromatin, the core histones not only create a structural backbone but also are crucial in regulating DNA replication, DNA repair, and transcription. In most multicellular eukaryotes, the H3 family contains three major isoforms, called H3 variants. The H3 barcode hypothesis proposed that the first layer of epigenetic information is provided by a specific pattern of H3 variants that are in turn recognized by specific reader proteins (*1*). Although intriguing, there has been little experimental evidence to support this hypothesis until now. On page 1281 of this issue, Davarinejad *et al.* (*2*) identify the first reader for the replication-coupled H3.1 variant. This reader initiates repair of DNA breaks encountered by the replication fork.

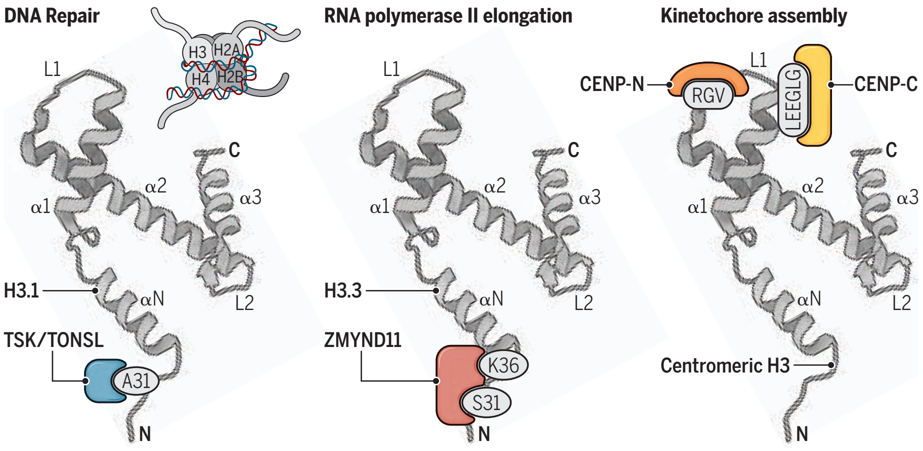
Histone variants are isoforms of core histones that acquired similar properties and functions through convergent evolution (*3*). They are found primarily among the H3, H2A, and H2B families. Certain variants have specific residues, and modification of these residues is critical for the variants' function. For example, in many eukaryotes, a serine specific to the H2A.X variant is phosphorylated at sites of DNA damage. This modification recruits a specific reader that in turn recruits DNA repair machinery (*4*). Conversely, variants can lack specific residues that are modified in the core histones. For example, there are flowering plant–specific H3 variants that either are not amenable to Lys^{27} trimethylation (K27me3) or lack the K27 residue altogether. In sperm, these variants remove the repressive K27me3 mark, enabling the activation of genes that are required for sperm function (*5*). It is thus clear that histone variants

Gregor Mendel Institute, Austrian Academy of Sciences, Vienna BioCenter, Vienna, Austria. Email: frederic.berger@gmi.oeaw.ac.at

GRAPHIC: N. CARY/SCIENCE

Readers of histone H3 variants

Specific proteins, or readers, bind to amino acid residues or sequences that are distinct to each H3 variant. This differential reader binding allows execution of a particular function.



CENP, centromere protein; TONSL, tonsoku-like; TSK, TONSOKU; ZMYND11, zinc finger MYND domain-containing protein 11.

can act upstream of the deposition of epigenetic marks.

Less well characterized are proteins that modify only specific histone variants. In several groups of eukaryotes, DNA replication–coupled H3 variants evolved independently. These replicative variants, H3.1 and H3.2, are characterized by an alanine residue at position 31 (A31) (*3*). This position is occupied by a threonine or serine residue in the H3.3 variant. In plants, A31 restricts to H3.1 the monomethylation of K27 (K27me1) by the histone-lysine *N*-methyltransferases ATXR5 and ATXR6 (*6*). These redundant enzymes are only found in land plants. Besides a clear association with constitutive heterochromatin, the exact relationship between H3.1, H3K27me1, and abnormal rereplication of heterochromatin in *atxr5*;*atxr6* double mutants has remained unclear.

Davarinejad *et al.* focused on H3.1 deposition at the DNA replication fork to identify a reader that can recognize the distinguishing A31 residue in *Arabidopsis thaliana*. They found that the tetratricopeptide repeat (TPR) domain of the TONSOKU (TSK) protein binds the amino-terminal tail of replicative H3.1. Structural analyses support the conclusion that A31 is essential for binding TSK and its ortholog in animals. *A. thaliana tsk* mutants were first characterized because of their enlarged meristems (*7*–*9*). They also showed high sensitivity to DNA-damaging agents and a genetic interaction with mutants in the chromatin-associated factor 1 complex, which deposits H3.1 at the replication fork (*9*), suggesting a link between H3.1 and TSK.

In animals, the TSK ortholog tonsoku-like (TONSL) was identified as an interactor of methyl methanesulfonate-sensitivity protein 22-like (MMS22L), a homolog of the eponymous yeast protein that is required for homologous recombination (HR) (*10*). HR is a major pathway for repair of DNA breaks during replication. TONSL is required for HR when the replication fork encounters a DNA double-strand break. Both TONSL and TSK recruit the DNA polymerase theta-associated end-joining machinery, which mediates the initial steps of HR (*10*). TONSL has an additional ankyrin repeat domain that binds H4 but is inhibited by H4K20me1, which is deposited behind the replication fork and thus marks replicated nucleosomes. As a result, TONSL is recruited only to the DNA replication fork (*11*). This ankyrin repeat domain is not present in TSK, and the H4K20me1 methyltransferase has not been identified in plants. Therefore, Davarinejad *et al.* suspected that TSK binding to H3.1 would be regulated through a distinct mechanism and found that H3K27me1 prevents TSK binding to H3.1 nucleosomes.

> **"…the first layer of epigenetic information is provided by a specific pattern of [histone] H3 variants that are in turn recognized by specific reader proteins."**

This study provides a compelling explanation for the genome instability and overreplication of heterochromatin observed in *atxr5*;*atxr6* double mutants, which has remained enigmatic since its description. Heterochromatin amplification in *atxr5*;*atxr6* double mutants is rescued in mutants defective in Rad17 and DNA polymerase theta, which create large tandem duplications (*12*). Davarinejad *et al.* propose that, in the absence of H3K27me1, TSK is recruited to regions away from replication forks, where it initiates HR, thereby causing the rereplication phenotype of *atxr5*;*atxr6* double mutants.

A parallel can thus be drawn between the role of H3K27me1 in plants and H4K20me1 in mammals in preventing spurious initiation of HR. Yet H3K27me1 confines the action of the TSK pathway to heterochromatin (in an unknown manner), whereas H4K20me1 is present in euchromatin. The biological importance of targeting these pathways to two different chromatin domains remains unclear. The parallel evolution of TONSL and MMS22L with H3 variants in eukaryotes is also intriguing. TSK is absent in yeast, which also lacks H3.1. Could selection of TSK be dependent on the presence of H3.1?

The study of Davarinejad *et al.* reports the first reader of replicative H3, adding to the very short list of readers of H3 variants—the kinetochore components centromere protein C (CENP-C) and CENP-N, which recognizes the centromeric H3 variant to establish the kinetochore during mitosis (*13*), and a protein specifically binding to H3K36me3 of H3.3 to influence RNA polymerase II elongation (*14*). These reader-variant pairs are now providing strong support to the H3 variant code hypothesis (*1*) (see the figure). It is foreseeable that in addition to the readers of phosphorylated H2A.X (*15*), new readers of H2A and H2B variants will be added to the list, expanding the idea that histone variants provide the foundation of chromatin functional organization. ■

REFERENCES AND NOTES

1. S. B. Hake, C. D. Allis, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 6428 (2006).
2. H. Davarinejad *et al.*, *Science* **375**, 1281 (2022).
3. B. Loppin, F. Berger, *Annu. Rev. Genet.* **54**, 121 (2020).
4. M. Papamichos-Chronakis, C. L. Peterson, *Nat. Rev. Genet.* **14**, 62 (2013).
5. M. Borg *et al.*, *Nat. Cell Biol.* **22**, 621 (2020).
6. Y. Jacob *et al.*, *Nature* **466**, 987 (2010).
7. S. Guyomarc'h, T. Vernoux, J. Traas, D.-X. Zhou, M. Delarue, *J. Exp. Bot.* **55**, 673 (2004).
8. T. Suzuki *et al.*, *Plant J.* **38**, 673 (2004).
9. S. Takeda *et al.*, *Genes Dev.* **18**, 782 (2004).
10. E. Duro *et al.*, *Mol. Cell* **40**, 632 (2010).
11. G. Saredi *et al.*, *Nature* **534**, 714 (2016).
12. J. A. Kamp, R. van Schendel, I. W. Dilweg, M. Tijsterman, *Nat. Commun.* **11**, 3615 (2020).
13. S. Chittori *et al.*, *Science* **359**, 339 (2018).
14. H. Wen *et al.*, *Nature* **508**, 263 (2014).
15. M. Stucki *et al.*, *Cell* **123**, 1213 (2005).

ACKNOWLEDGMENTS

I thank J. M. Watson for proofreading. This work was supported by the Gregor Mendel Institute.

10.1126/science.abo4219

GRAPHIC: A. MASTIN/*SCIENCE*

ORGANIC CHEMISTRY

Radical coupling decreases synthetic burden

Cross coupling of light-promoted β-keto radicals enables natural product syntheses

By **Spencer P. Pitre**

Natural products play a prominent role in the discovery of new pharmaceutics and medical therapies (*1*), given their rich chemical diversity that is optimized for interactions with biological macromolecules (*2*). However, isolation of sufficiently large quantities of these complex molecules often proves too costly or unsustainable, making natural product total synthesis an area of importance for organic chemists. Constructing these complex molecular scaffolds often necessitates development of new reactions, and such studies often provide valuable insights into functional group compatibility and a more complete understanding of reaction scope. On page 1270 of this issue, Landwehr *et al.* (*3*) report a light-promoted radical cross-coupling reaction to simplify access to neuroactive metabolites extracted from the rainforest canopy trees *Galbulimima belgraveana* (GB) (*4*). Merging two partially aromatic building blocks generated a key intermediate that could be efficiently reduced globally, selectively, or both to access three distinctive natural products.

To gain a deeper understanding of the biological targets of these neuroactive metabolites, a more concise and general strategy to access these natural products was needed. Landwehr *et al.* envisioned that GB22 was poised for further diversification through selective reduction of its aromatic ring. Through retrosynthetic analysis, a key aromatic intermediate was identified that could lead to the much more complex GB22 in only three successive steps. However, approaches to merge the two aromatic subunits that rely on traditional, two-electron reactivity, such as enone conjugate addition and Friedel-Crafts addition, proved unfruitful.

To circumvent these issues, two aromatic partners were united through the trapping and coupling of a β-keto carbon–centered radical. For many years, reactions proceeding through the intermediacy of carbon-centered radicals received little attention in the synthesis of complex molecules. The techniques for generating free radicals often led to unpredictable reactions that formed complex mixtures of products. However, the last two decades have witnessed a dramatic increase in the use of free-radical reactions, and carbon-centered radicals have become important reactive intermediates in the synthesis of structurally elaborate molecules (*5*, *6*).

New strategies for more controlled radical generation include, but are not limited to, photochemical redox reactions initiated by visible light that generate free radicals under mild and more easily controlled conditions compared with those of thermal methods (*7*, *8*). Although much of the same reactivity that can be attained with photoredox chemistry can be accomplished by using reagents that are either strong oxidants or reductants, these reagents are inherently difficult to handle because of their high reactivity. In photoredox reactions, a photosensitizer (or photocatalyst) generates potent single-electron reductants and oxidants in situ (*9*). The ability for these photocatalysts to absorb visible light is key to the success of photoredox reactions because most organic molecules do not absorb in this region, and product decomposition and side reactions can be avoided.

In 2014, Zuo *et al.* (*10*) and Tellis *et al.* (*11*) reported the merger of a photoredox catalytic cycle with a nickel catalytic cycle to yield cross-coupled products that were previously inaccessible. Landwehr *et al.* used an organic photocatalyst that oxidatively cleaved a siloxypropane moiety on one of the aromatic partners to generate a β-keto carbon–centered radical (*12*) that could be trapped by a nickel catalyst and cross-coupled with an aryl halide to yield the key aromatic intermediate (see the figure). They selectively reduced the aromatic rings of this simple aromatic intermediate to rapidly introduce the needed molecular complexity to construct three distinct congeners—(±)-GB22, (±)-GB13, and (±)-himgaline—in only six to nine synthetic steps, or nearly one-third of the prior best attempts.

These efficient syntheses and the ease of diversification should provide a means to deconvolute the biological targets, functions, and translational potential of the GB natural products. Landwehr *et al.* simplified access to these metabolites through rapid introduction of molecular complexity through global or selective reductions of low-complexity intermediates, a strategy that could be explored with other natural products. ■

Department of Chemistry, Oklahoma State University,
Stillwater, OK 74078, USA. Email: spencer.p.pitre@okstate.edu

REFERENCES AND NOTES

1. A. G. Atanasov, S. B. Zotchev, V. M. Dirsch, C. T. Supuran, *Nat. Rev. Drug Discov.* **20**, 200 (2021).
2. J. Hong, *Curr. Opin. Chem. Biol.* **15**, 350 (2011).
3. E. M. Landwehr *et al.*, *Science* **375**, 1270 (2022).
4. S. V. Binns *et al.*, *Aust. J. Chem.* **18**, 569 (1965).
5. S. P. Pitre *et al.*, *J. Am. Chem. Soc.* **141**, 2800 (2019).
6. K. J. Romero, M. S. Galliher, D. A. Pratt, C. R. J. Stephenson, *Chem. Soc. Rev.* **47**, 7851 (2018).
7. S. P. Pitre, L. E. Overman, *Chem. Rev.* **122**, 1717 (2022).
8. T. P. Nicholls, D. Leonori, A. C. Bissember, *Nat. Prod. Rep.* **33**, 1248 (2016).
9. N. J. Turro, V. Ramamurthy, J. C. Scaiano, *Principles of Molecular Photochemistry: An Introduction* (University Science Books, 2009).
10. Z. Zuo *et al.*, *Science* **345**, 437 (2014).
11. J. C. Tellis, D. N. Primer, G. A. Molander, *Science* **345**, 433 (2014).
12. H. Rinderhagen, P. A. Waske, J. Mattay, *Tetrahedron* **62**, 6589 (2006).

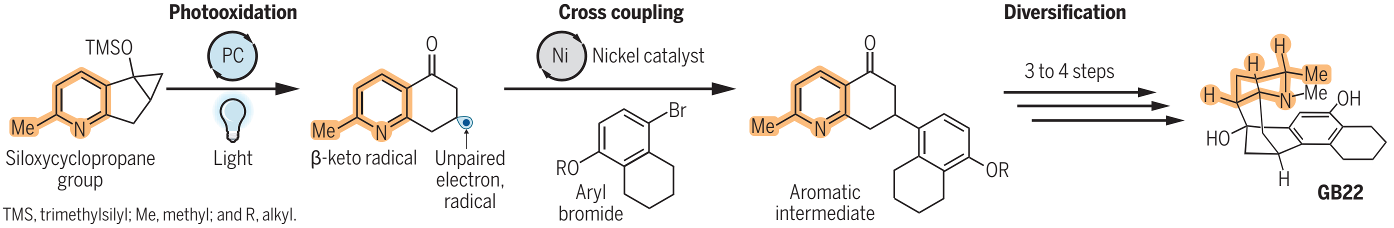
ACKNOWLEDGMENTS

I thank N. A. Weires for his insight and feedback.

10.1126/science.abo2398

Light-promoted radical reactions

Landwehr *et al.* used a photocatalyst (PC) to generate a β-keto carbon–centered radical in the synthesis of natural products from the *Galbulimima belgraveana* (GB) family.



TMS, trimethylsilyl; Me, methyl; and R, alkyl.

GRAPHIC: N. CARY/SCIENCE

RETROSPECTIVE

Luc Montagnier (1932–2022)

Discoverer of the human immunodeficiency virus

By **Robin A. Weiss**[1] and **Simon Wain-Hobson**[2]

Luc Montagnier died on 8 February at the age of 89. In 1983, he led the team of investigators at the Institut Pasteur in Paris that described a novel virus now named human immunodeficiency virus (HIV), which causes acquired immunodeficiency syndrome (AIDS). In 2008, Montagnier and his colleague Françoise Barré-Sinoussi were awarded one-half of the Nobel Prize in Physiology or Medicine for identifying HIV.

Montagnier was born in Chabris, France, on 18 August 1932. His father engaged in amateur scientific experiments, which inspired his son to become a scientist. He studied medicine at the University of Poitiers, graduated from the University of Paris–Sorbonne in 1960, and joined the Centre national de la recherche scientifique (CNRS). He married Dorothea Ackerman in 1961, with whom he had three children. He came to the UK to work with F. Kingsley Sanders on RNA viruses, and then with Ian Macpherson, with whom he showed that polyomavirus-transformed cells, unlike their normal counterparts, could grow into tumor cell colonies suspended in soft agar culture. Returning to France in 1964, Montagnier worked at the Institut Curie until 1972, when Jacques Monod invited him to work at the Institut Pasteur. Joined in 1975 by Barré-Sinoussi and Jean-Claude Chermann, he stayed there until his retirement in 2000.

AIDS was first observed in gay men in 1981. By 1982 it became apparent that AIDS also occurred in injecting drug users and in recipients of blood transfusions and blood products. The hunt was on to identify an infectious cause. Amid speculation about the nature of the agent, American biomedical researcher Robert Gallo championed retroviruses related to the human T cell leukemia virus that he had discovered in 1980.

Montagnier's work on AIDS started when physician Willy Rozenbaum provided him with a lymph node biopsy taken from a gay man with persistent lymphadenopathy, a sign of incipient AIDS. Montagnier stimulated T cells from the biopsy to proliferate in vitro. The cultures grew poorly and were passed to Barré-Sinoussi in Chermann's laboratory; she detected reverse transcriptase activity (the hallmark of a retrovirus) in the culture medium. The culture could only be maintained by adding fresh lymphocytes. Electron microscopy revealed retrovirus-like particles budding from infected cells.

The isolation of a cytopathic retrovirus from a single patient did not make headlines when it was published in May 1983. Montagnier's team, however, soon reported further isolates from AIDS patients, including a heterosexual couple from Africa and later from hemophiliacs. Key clinical researchers including Françoise Brun-Vézinet and Christine Rouzioux contributed samples and were instrumental in developing serological assays for antibodies to HIV, which enabled diagnostic tests for blood screening. Electron microscopist Charles Dauguet showed that the virus particles closely resembled those of a lentivirus, while immunologist David Klatzmann showed that the virus had a specific tropism for CD4$^+$ T cells.

PHOTO: FRANCOIS LOCHON/GAMMA-RAPHO/GETTY IMAGES

Molecular biologists Marc Alizon and Pierre Sonigo, led by one of us (S.W.-H.), cloned and sequenced the HIV genome, demonstrating the features of a lentivirus. Montagnier's group was also the first to report a different AIDS-related retrovirus in West Africa, later called HIV-2. These multidisciplinary discoveries were essential for our early understanding of HIV. Confirmation that a novel retrovirus really was the culprit behind AIDS was cemented in 1984 by reports from Gallo's team and groups in Atlanta and San Francisco. The identity of Gallo's isolate with the French virus was not immediately apparent, which led to a heated patent dispute, ultimately resolved in 1991 with the realization that the two isolates were but one, originating in Paris.

Initially, Montagnier, Chermann, and Barré-Sinoussi were regarded as a trio, but the camaraderie did not last. By the time a new laboratory building for AIDS research was constructed on the Pasteur campus, Chermann had departed for Marseille, and Barré-Sinoussi decided to remain in her cramped laboratory across campus. Nobelist François Jacob, who never highly rated Montagnier, dubbed the new building "La Basilique St. Luc."

Montagnier was a reserved character who seldom seemed able to inspire or praise his collaborators. He was not a natural communicator, despite his good grasp of English, and was ill-equipped to handle his rival Gallo's flamboyant manner. Sadly, as his fame increased, he became increasingly drawn to dubious theories, including his conviction that mycoplasma was an essential cofactor in AIDS and that DNA emitted electromagnetic radiation, among other fringe ideas. His grandiose research foundations in New York and Africa achieved little.

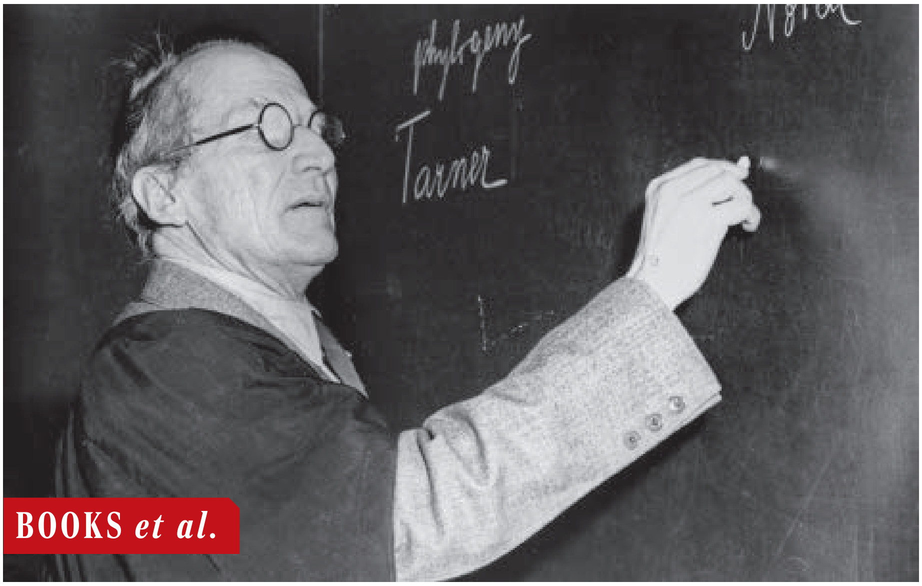
Nonetheless, the discovery of HIV changed many careers, including ours. I (S.W.-H.) initiated the molecular analysis of HIV while still in Pierre Tiollais's hepatitis laboratory. My collaboration with Montagnier led to many joint studies, and I was soon appointed to a faculty position in his department. I (R.A.W.) was hesitant to become involved in HIV after the initial French paper until Montagnier showed me more convincing unpublished electron micrographs and data on the new isolates from AIDS patients. He kindly provided samples of HIV, enabling my laboratory to investigate neutralizing antibodies and mechanisms of HIV entry into cells, and we coauthored some papers. In 1992, I spent a brief sabbatical at Institut Pasteur and enjoyed chatting with Montagnier over coffee.

Montagnier received many awards for his work on HIV prior to the Nobel Prize, notably the Albert Lasker Clinical Medical Research Award and the Louis-Jeantet Prize. He was promoted to grand officier de la Légion d'honneur in 2008. Many universities across the globe awarded him honorary doctorates. The discovery of HIV led to radical advances in virology and immunology. With antiretroviral drug treatment, HIV infection is now a largely manageable condition, despite the lack of an efficacious vaccine. As Björn Vennström remarked at the 2008 Nobel ceremony, Montagnier was in the right place at the right time. ■

10.1126/science.abo7630

---

[1] Division of Infection and Immunity, University College London, London, UK. [2] Institut Pasteur, Paris, France. Email: r.weiss@ucl.ac.uk; simon.wain-hobson@pasteur.fr



BOOKS *et al.*

Schrödinger lectures at a blackboard, circa 1950.

SCIENCE LIVES

Schrödinger's Oxford years

A welcome new tome offers an incomplete portrait of a flawed genius

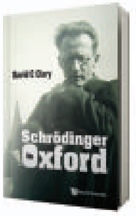
By **Cormac O'Raifeartaigh**

In a new biography, Sir David Clary—professor of chemistry at the University of Oxford and former president of Magdalen College—presents a meticulous account of the life and work of Erwin Schrödinger, with an emphasis on the famous physicist's time at Oxford. The book provides many new insights into the career of the brilliant Austrian scientist. However, it includes very little consideration of Schrödinger's character or personal life, an unfortunate omission given the recent resurgence of allegations concerning Schrödinger's behavior toward young women.

Many readers will be aware of previous Schrödinger biographies, ranging from Walter Moore's definitive *Schrödinger: Life and Thought* to John Gribbin's highly readable *Erwin Schrödinger and the Quantum Revolution*. However, neither of the aforementioned books gives a comprehensive account of Schrödinger's years at Oxford. Clary's book is therefore a welcome addition to this literature.

Much of the material covered early in the book describing the scientific context in which Schrödinger developed his famous equation has appeared elsewhere. Yet Clary's account makes for fascinating reading, not least because of his clear style and copious citation of primary sources and original scientific articles. Indeed, the author's depiction of Schrödinger's great breakthrough in the context of the work of quantum pioneers such as Louis de Broglie, Albert Einstein, Niels Bohr, and Werner Heisenberg is exemplary history of science writing.

Similarly, Clary's description of Schrödinger's departure in 1933 from a highly eminent position at the University of Berlin to a precarious, untenured position at Magdalen College at Oxford makes for fascinating reading. The story of Schrödinger's flight from a threatening political situation in Germany to the leafy cloisters of Oxford, followed by a foolish and short-lived move to Austria, is also well known, but the author provides a compelling narrative of this episode with political and scientific considerations deftly woven together.



Schrödinger in Oxford
David C. Clary
World Scientific, 2022.
420 pp.

The central part of the book, in which Clary gives a detailed description of Schrödinger's time at Magdalen College, will probably be the section of most interest to many readers. Schrödinger, like Einstein before him, was unhappy at Oxford. With copious reference to letters and other original sources, Clary examines the reasons for this unhappiness in detail, which included sociological factors (Schrödinger disliked the college's pomp and formality) as well as academic considerations (such as the lack of a community of brilliant theoretical physicists similar to that of Berlin and Zurich). The author also explores the disruptive effect Schrödinger's search for a permanent academic position had on his work in these years, an important consideration often overlooked in other biographies.

However, Clary's painstaking presentation of excerpts from letters offering employment to Schrödinger does not make for light reading. Indeed, the author has a marked tendency to present the fruits of his meticulous research in the main text rather than in footnotes or references, at times complicating the narrative unnecessarily.

Likewise, in the section describing the award of the 1933 Nobel Prize in Physics to Schrödinger and Paul Dirac, it is interesting (but not that surprising) to learn that Schrödinger received many letters of congratulation from his colleagues. One wonders, however, why the author felt the need to display excerpts from many of these letters. Such passages might have been replaced by a deeper consideration of Schrödinger's work in quantum theory while at Oxford. Indeed, there is surprisingly little discussion of the physicist's philosophical difficulties with the theory he helped create—or of the nature of his reservations concerning the conventional interpretation of his famous wave function.

In the book's final chapter, Clary indicates that he has intentionally avoided any discussion of Schrödinger's "extra-marital affairs," pointing out that "he is not the only famous theoretical physicist to have a complicated personal life." This is a legitimate approach to biography, but it is an unfortunate omission given the recent resurgence of allegations that, during his time at the Dublin Institute for Advanced Studies, Schrödinger may have pursued a predatory interest in female adolescents (*1*). Indeed, there are several well-known instances where the famous scientist appears to have groomed young girls for future relationships. As recently pointed out in *The Irish Times*, these findings bear scrutiny and complicate Schrödinger's legacy (*2*). ■

The reviewer lectures in physics at Waterford Institute of Technology, Waterford, Ireland, and is a fellow of the Institute of Physics and an adjunct research fellow at the Dublin Institute for Advanced Studies, Dublin, Ireland. Email: coraifeartaigh@wit.ie

REFERENCES AND NOTES

1. W. Moore, *Schrödinger: Life and Thought* (Cambridge Univ. Press, 1989).
2. J. Humphreys, "How Erwin Schrödinger indulged his 'Lolita complex' in Ireland," *The Irish Times*, 11 December 2021.

10.1126/science.abo1777

PHOTO: SSPL/GETTY IMAGES

SCIENCE LIVES

Drawing the mind, one neuron at a time

A new biography celebrates the Spaniard who founded modern neuroscience

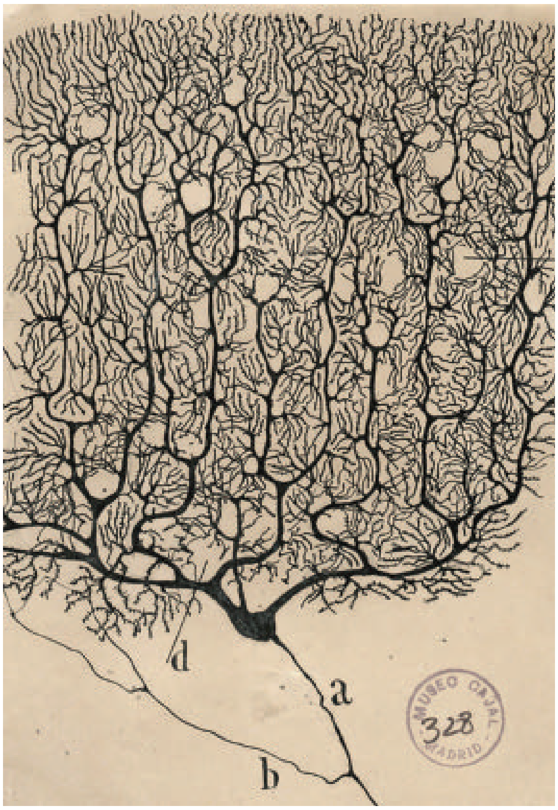
By **Alex Gomez-Marin**

In his new book, *The Brain in Search of Itself: Santiago Ramón y Cajal and the Story of the Neuron*, author Benjamin Ehrlich walks readers through the life and work of Santiago Ramón y Cajal, the Spaniard whom many consider the founder of modern neuroscience. Structured as a series of brief, captivating chapters, the first major English language biography of Cajal is a delightful read. With rigor and eloquent prose, Ehrlich captures the essence of Cajal's visionary thought and force of character, drawing a composite portrait of the "peasant genius" who became a Nobel laureate.

Born in a small village in the north of Spain in 1852, Cajal had a difficult upbringing. As a child, he was beaten by his father, Justo, at home and by his teachers at school, a terror that he later admitted he would "prefer to bury in the shadows of the unconscious." At the age of 10, he knew he wanted to be a professional artist, but Justo ordered him to become a doctor first. Cajal thus "entered the Castle of Science through the Door of Art."

As a university student, Cajal loved writing stories. In one, he arrives on Jupiter, which is inhabited by beings so big that he can inspect their cells directly, foreshadowing his scientific contributions. With his first microscope, he began a systematic exploration of the tissues of the human body. Not content to superficially observe cells, he sought to chart their microscopic anatomy and turned to histology, the art of optical dissection.

In 1879 he married Silveria Fañanás, about whom he noted: "Only the unsurpassable self-abnegation of my wife made my scientific work possible." Together, they had seven children. The Cajals' modest economy covered their rent and food, with the rest spent on the scientist's home laboratory. When he finally won his third "oposición"—a "quintessentially Spanish" tenure examination—they moved from Zaragoza to Valencia.



Cajal's rendering of a Purkinje neuron found in the human cerebellum captures the beauty and intricacy of this complex cell.

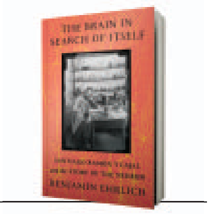
ILLUSTRATION: SANTIAGO RAMÓN Y CAJAL/WIKIMEDIA COMMONS

From Luis Simarro Lacabra, Cajal's colleague and competitor in Madrid, he learned about Camillo Golgi's staining technique that made visible random bits of nervous tissue. Working in solitude and employing his prodigious imagination, he avidly used and improved the method, even as most others abandoned it as too unreliable.

Weary of distractions, Cajal moved to Barcelona at the end of 1887, away from family and friends. In his first 4 years there, he published 43 papers and founded his own journal. Instead of trying to map the impenetrable "full grown forest" of the adult brain, Cajal shifted his studies to the "young wood" of early embryos. Absent the sheaths of myelin that coated mature neurons, young neurons stained better, he found.

In the cerebellum, he found magnificent Purkinje cells, the largest of neurons. Whereas Golgi had drawn them fading away, in Cajal's eyes they "ended freely," and his empirical data compellingly suggested the independence of each neuron. He also inferred that electrical transmission in neurons was directional. To communicate his seminal discoveries, Cajal traveled to Berlin in 1889, where he met the renowned anatomist Albert von Kölliker, who began promoting the Spaniard's work to the world.

The Brain in Search of Itself: Santiago Ramón y Cajal and the Story of the Neuron
Benjamin Ehrlich
Farrar, Straus and Giroux, 2022. 464 pp.



After Cajal's landmark discoveries in Barcelona, he moved to Madrid in 1892, where prizes and honors poured in and his "Spanish school," which teemed with promising young scientists, started taking off. He published dozens of monographs, including his masterpiece *Texture of the Nervous System of Man and the Vertebrates*, an overwhelming synthesis comprising 2000 pages and close to 1000 illustrations.

In 1906, he received the Nobel Prize in Physiology or Medicine for establishing that the neuron is the basic unit of the nervous system, sharing the award with Golgi. While the Italian provided the key to unlock the secrets of the nervous system, the Spaniard mastered its use. Urged to take part in politics by his liberal colleagues, Cajal mostly demurred, dedicating his scarce free time to establishing Spanish research institutions.

Cajal spent his final days surrounded by his "spiritual grandchildren"—disciples Jorge Francisco Tello and Fernando de Castro, among others—passing away in 1934 at the age of 82. He requested no solemn homage. "And what do praises matter to me?" he wrote once in a notebook. "When they applaud me I will not exist."

Yet it is important to remember figures like Cajal, not only to honor them but also because their stories help us better understand ourselves. With this delightful biography, Ehrlich has done justice to Cajal, just as Cajal did justice to the brain. ■

10.1126/science.abo0190

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

Brazil's caves, home to diverse species and minerals, were stripped of protections by a recent presidential decree.

PHOTO: DANIEL MENIN

Edited by **Jennifer Sills**

Brazilian cave heritage under siege

Under President Jair Bolsonaro, Brazil has neglected environmental protection (*1*). Until recently, subterranean environments were largely safeguarded; although some caves could be destroyed for the purposes of exploiting mineral resources or urban development, those classified as presenting maximum cultural, geological, and/or biological value had to be preserved (*2*). However, a January presidential decree (*3*) allows the destruction of even those caves with maximum relevance, representing a substantial step backward for subterranean conservation.

The new decree ignores the irreplaceable scientific value of Brazilian caves. Researchers have documented hundreds of endemic obligate cave species, including animals with unique adaptations (*4*). Thousands of species still lack formal descriptions, and many more species await discovery. At least 72 bat species roost in Brazilian caves and provide critical ecosystem services such as pest control (*5*). The geodiversity of these caves is also exceptional, with many sites harboring rare minerals and geological formations found nowhere else (*6*). Finally, the multitude of archeological and paleontological cave sites across the country provide a unique record of Brazil's past.

Bolsonaro's decree increases the extinction risk of unique species. By ignoring the intrinsic and utilitarian values of Brazilian caves, the policy neglects global conservation strategies to safeguard the subterranean biome (*7*, *8*), violates the principles of both the Federal Biodiversity Policy (*9*) and the Convention on Biological Diversity (*10*), and fails to align with Sustainable Development Goals. The potential losses of unique species, ecosystem services, and new industrial chemicals (which could be derived from the microbes found in the caves) (*11*) are unpredictable. In addition, because causing species extinction could damage a company's reputation, the policy could undermine the mining sector that the administration seeks to support. We cannot allow these ecosystems to be obliterated by short-sighted decrees.

Rodrigo Lopes Ferreira[1]*, Enrico Bernard[2], Francisco William da Cruz Júnior[3], Luis Beethoven Piló[2], Allan Calux[4], Marconi Souza-Silva[1], Jos Barlow[5], Paulo S. Pompeu[6], Pedro Cardoso[7], Stefano Mammola[7,8], and 85 additional authors†

[1]Centro de Estudos em Biologia Subterrânea, Departamento de Ecologia e Conservação, Universidade Federal de Lavras, Lavras, MG 37200-900, Brazil. [2]Laboratório de Ciência Aplicada à Conservação da Biodiversidade, Departamento de Zoologia, Universidade Federal de Pernambuco, Recife, PE 50670-901, Brazil. [3]Instituto de Geociências, Universidade de São Paulo, São Paulo, SP 05508-080, Brazil. [4]Carstografica, Karst Applied Research Centre, Belo Horizonte, MG 31170-320, Brazil. [5]Lancaster Environment Centre, Lancaster University, Lancaster, UK. [6]Departamento de Ecologia e Conservação, Universidade Federal de Lavras, Lavras, MG, Brazil. [7]Laboratory for Integrative Biodiversity Research, Finnish Museum of Natural History (Luomus), University of Helsinki, Helsinki, Finland. [8]Molecular Ecology Group, Water Research Institute, National Research Council (CNR-IRSA), Verbania Pallanza, Italy.

*Corresponding author. Email: drops@ufla.br

†Full list of authors and affiliations can be found at www.science.org/doi/10.1126/science.abo1973.

REFERENCES AND NOTES

1. H. Escobar, *Science*, 10.1126/science.aay9103 (2019).
2. Decreto nº 6.640, de 07 de novembro de 2008 (2008); www.planalto.gov.br/ccivil_03/_Ato2007-2010/2008/Decreto/D6640.htm [in Portuguese].
3. Decreto nº 10.935, de 12 de janeiro de 2022 (2022); www.in.gov.br/en/web/dou/-/decreto-n-10.935-de-12-de-janeiro-de-2022-373591582 [in Portuguese].
4. K. Yoshizawa *et al.*, *Curr. Biol.* **24**, 9 (2014).
5. R. A. Medellín, R. Wiederholt, L. Lopez-Hoffman, *Biol. Conserv.* **211**, 45 (2017).
6. C. J. Schenk, R. J. Viger, C. P. Anderson, "Maps showing geology, oil, and gas fields and geologic provinces of the South America region: U.S. Geological Survey Open-File Report 97-470-D" (1999); https://pubs.er.usgs.gov/publication/ofr97470D.
7. J. J. Wynne *et al.*, *Conserv. Lett.* **14**, e12834 (2021).
8. D. Sánchez-Fernández *et al.*, *Nat. Clim. Change* **11**, 458 (2021).
9. Decreto nº 4.339 de 22 de agosto de 2002 (2002); www.planalto.gov.br/ccivil_03/decreto/2002/D4339.htm [in Portuguese].

10. Convention on Biological Diversity (www.cbd.int).
11. P. N. da Costa Souza *et al.*, *Afr. J. Microbiol. Res.* **7**, 483 (2013).

10.1126/science.abo1973

Brazil's mangroves: Natural carbon storage

Brazil has the second-largest area of mangrove cover in the world (*1*), including the world's largest continuous mangroves (*2*). These ecosystems provide food security for coastal populations; habitat for terrestrial, bird, and fish species; and carbon sequestration (*3*). However, Brazil's mangroves are in peril. In 2020, the Brazilian government approved legislation that would have eliminated mangrove protection for the benefit of real estate development (*4*). Fortunately, in December 2021, the Brazilian Supreme Court found the controversial policy unconstitutional (*5*). Brazil must continue to protect its mangroves despite an administration that has shown disregard for the environment (*6*).

Mangrove conservation doubles as an effective natural climate solution (*7*). Mangroves can store up to 10 times more carbon per hectare than upland forests (*8*), and Brazil's mangrove soil currently stores about 3 to 8% of global carbon (*9*). Yet these ecosystems and their vast natural carbon storage potential are vulnerable to emissions from anthropogenic activities and, if lost, could not be restored by 2050 (*10*).

The attempts to change legislation to prioritize development over mangroves demonstrate that current protections are threatened. Including Brazil's mangroves in the national Reducing Emissions from Deforestation and Forest Degradation (REDD) strategy could be a way to maintain the carbon stocks and better protect their biodiversity. Although there is a civil society initiative for monitoring Brazilian mangroves (*11*, *12*), Brazil should also create an official monitoring program to ensure effective conservation and enforcement of the policies in place to protect these ecosystems.

Denilson da S. Bezerra[1]*, Adriano de Lima Santos[1], Janaina Santos Bezerra[1], Silvana Amaral[2], Milton Kampel[2], Liana O. Anderson[3], Flávia Rebelo Mochel[1], Jorge Luiz Silva Nunes[1], Naíla Arraes de Araujo[1], Larissa Nascimento Barreto[1], Maria do S. S. Pinheiro[1], Marcio José Celeri[1], Fabrício B. Silva[4], Alexsandro Mendonça Viegas[5], Stella Manes[6], Taissa C. S. Rodrigues[7], Josué C. Viegas[8], Ulisses D. V. Souza[9], André L. S. Santos[10], Celso H. L. Silva-Junior[2,11]

[1]Universidade Federal do Maranhão, São Luís, Maranhão, Brazil. [2]Instituto Nacional de Pesquisas Espaciais, São José dos Campos, São Paulo, Brazil. [3]Centro de Monitoramento e Alertas a Dessastres Naturais, São José dos Campos, São Paulo, Brazil. [4]Universidade Ceuma, São Luís, Maranhão, Brazil. [5]Escola de Governo do Maranhão, São Luís, Maranhão, Brazil. [6]Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil. [7]Universidade Estadual da Região Tocantina do Maranhão, Imperatriz, Maranhão, Brazil. [8]Universidade de Coimbra, Coimbra, Portugal. [9]Colégio Universitário da Universidade Federal do Maranhão, São Luís, Maranhão, Brazil. [10]Instituto Federal de Educação, Ciência e Tecnologia do Maranhão, São Luís, Maranhão, Brazil. [11]Universidade Estadual do Maranhão, São Luís, Maranhão, Brazil.
*Corresponding author.
Email: denilson.bezerra@ufma.br

REFERENCES AND NOTES

1. M. Spalding, *World Atlas of Mangroves* (Routledge, 2010).
2. W. R. Nascimento *et al.*, *Estuar. Coast. Shelf Sci.* **117**, 83 (2013).
3. S. Y. Lee *et al.*, *Glob. Ecol. Biogeogr.* **23**, 726 (2014).
4. A. C. Amaral, "Salles atacará normas de proteção de terras, águas e ar em revogaço nesta segunda" (2020); https://ambiencia.blogfolha.uol.com.br/2020/09/27/salles-atacara-normas-de-protecao-de-terras-aguas-e-ar-em-revogaco-nesta-segunda [in Portuguese].
5. P. Watanabe, "STF restaura proteção a mangues e restingas, por unanimidade" (2021); https://www1.folha.uol.com.br/ambiente/2021/12/stf-restaura-protecao-a-mangues-e-restingas-por-unanimidade.shtml [in Portuguese].
6. F. P. Ottoni *et al.*, *Biota Neotrop.* **21**, 2 (2021).
7. P. I. Macreadie *et al.*, *Nat. Rev. Earth Environ.* **2**, 826 (2021).
8. J. Howard *et al.*, *Front. Ecol. Environ.* **15**, 42 (2017).
9. V. Hatje *et al.*, *Limnol. Oceanogr.* **66**, 321 (2021).
10. M. L. Noon *et al.*, *Nat. Sustain.* **5**, 37 (2022).
11. C. M. Souza *et al.*, *Remote Sens.* **12**, 2735 (2020).
12. C. Diniz *et al.*, *Remote Sens.* **11**, 808 (2019).

10.1126/science.abo4578

Disrupting targets' dependency on bullies

Targets of academic bullying sometimes despair about the lack of institutional support they receive. Even when universities take a firm stance against perpetrators [e.g., (*1*, *2*)], it is very difficult to disrupt the dependency of the target on the bully. Institutions must put policies in place to allow targets to extricate themselves and their work from the perpetrator without compromising their careers.

To address the issue of academic bullying, many recommendations focus on punishing perpetrators (e.g., prohibiting the known bullies from leading labs), which is in line with domestic abuse legislation in some countries, such as Austria (*3*). However, simply removing the bully does not protect the target from their influence. Another common solution is to transfer targets to another lab (*4*–*7*), but many targets do not perceive this as a viable option; after years trying to pursue their research goals, changing labs would mean losing their projects, associated publications, and access to letters of recommendation, which can be career-breaking for early-stage academics.

The scientific community must implement policies that grant targets of bullying independence. For example, if the allegations of bullying behaviors are validated (e.g., through confirmation by institutional investigation committees), institutions can grant the targets the rights to the lab's data, allowing them to publish. Institutions can also ban the perpetrators from submitting letters of recommendation for targets and designate others, such as department chairs, to write such letters instead. Medical students, PhD students, and postdocs should have the right to independently pursue research and publication projects when supervisors are known to engage in bullying behaviors. This would serve as a fail-safe for situations in which the career success of the target would otherwise be contingent on staying in a bully's group.

Granting bullying targets these rights would allow them to protect their mental health without sacrificing their professional success. If lab members knew that they could continue their work independently, they would likely tolerate fewer bullying behaviors. Empowerment would lower the bar to reporting inappropriate behavior as well. Protecting targets is crucial to stemming the tide of academic bullying behavior.

Susanne Täuber[1]* and Morteza Mahmoudi[2]

[1]Department of Human Resource Management & Organizational Behavior, University of Groningen, Groningen, Netherlands. [2]Department of Radiology and Precision Health Program, Michigan State University, East Lansing, MI 48824, USA.
*Corresponding author. Email: s.tauber@rug.nl

REFERENCES AND NOTES

1. C. Flaherty, "No more passing the harasser," *Inside Higher Education* (2018); www.insidehighered.com/news/2018/09/25/u-wisconsin-system-proceeds-plan-disclose-misconduct-findings-against-employees.
2. M. N. Bolotnikova, "Jorge Domínguez loses emeritus status after misconduct investigation," *Harvard Magazine* (2019). www.harvardmagazine.com/2019/05/jorge-domínguez-loses-emeritus-status.
3. B. Haller, "The Austrian legislation against domestic violence" (2005); www.ikf.ac.at/english/austrian_legislation_against_domestic_violence.pdf.
4. J. Stebbing *et al.*, *Postgrad. Med. J.* **80**, 93 (2004).
5. L. Yarwood-Ross, C. Haigh, *Nurse Res.* **24**, 38 (2014).
6. K. H. Hunter, K. Devine, *Int. J. Doct. Stud.* **11**, 35 (2016).
7. S. Moss, M. Mahmoudi, *EClinicalMedicine* **40**, 101121 (2021).

COMPETING INTERESTS

S.T. is a member of the advisory board at the Academic Parity Movement (www.paritymovement.org), a nonprofit organization dedicated to addressing academic discrimination, violence, and incivility. M.M. is a cofounder and director of the Academic Parity Movement and receives royalties/honoraria for his published books, plenary lectures, and licensed patents.

10.1126/science.abo3412


Where
Science
Gets
Social.


Share

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

WILDFIRES

Fired up

Large wildfires can produce ascending atmospheric plumes of such great intensity that they inject smoke and other combustion products into the stratosphere. Bernath *et al.* show that compounds transported into the stratosphere by the Black Summer Australian fires in 2019–2020 caused extreme perturbations in stratospheric gas composition that have the potential to destroy ozone. As climate change causes severe wildfires to become more frequent, their effects on the global ozone budget will grow. —HJS

Science, abm5611, this issue p. 1292

Wildfires in Australia in 2019–2020 released massive plumes of smoke and combustion products.

PHOTO: GLEN MOREY/AP IMAGES

HUMAN GENOMICS

From genetic variant to phenotypic trait

Quantitative trait loci (QTLs) are regions of the genome that are statistically associated with a specific phenotype. However, mapping QTLs to a causal genetic variant is difficult because most target loci are in noncoding parts of the genome and in regions containing multiple, closely spaced variants. To identify how genes link to a phenotype, Abell *et al*. applied a massively parallel reporter assay to targeted regions of the genome to determine the links between genetic variants and human phenotypes. The group identified causal variants for disease loci, often including transcription factor–binding sites. Causal variant effects were shown to be weaker than previously hypothesized because the genetic underpinning of a phenotype may be due to multiple proximal causal genetic variants. —LMZ

Science, abj5117, this issue p. 1247

ULTRACOLD CHEMISTRY

Layers of ultracold chemistry

Ultracold polar molecules trapped in two-dimensional layers are predicted to exhibit complex quantum phenomena not available with other platforms because of long-range anisotropic and tunable dipolar interactions. Using precision electric field control, Tobias *et al*. demonstrated layer-resolved creation and imaging of ultracold potassium–rubidium molecules confined to two-dimensional planes in an optical lattice. They also studied spin exchange and chemical reactions, which have been shown to be highly dependent on the molecule temperature and interlayer detunings introduced by an electric field gradient. This work demonstrates high control of ultracold molecules in an optical lattice and is a promising step toward exploring emerging phenomena in quantum gas systems with reduced dimensionality. —YS

Science, abn8525, this issue p. 1299

URBAN EVOLUTION

Plants adapt to city environments

Urban development alters the local environment, potentially driving rapid evolution. Santangelo *et al*. collected data on white clover populations from 160 cities to test for consistent responses to urban environments. They found that the production of an antiherbivore chemical defense increased with greater distance from the urban center in many cities. Genomic data suggest that this trend is adaptive, likely in response to lowered drought stress and herbivory pressure in urban centers. This study from the Global Urban Evolution

Project provides evidence of widespread adaptation to urbanization. —BEL

Science, abk0989, this issue p. 1275

METALLURGY

Pushing the grain boundaries

The behavior of grain boundaries in metals during deformation is important because it can dictate the macroscopic behavior. Lihua *et al*. used aberration-corrected in situ electron microscopy observation of platinum grain boundaries during straining to detail how they evolve. The authors observed grain boundary sliding, which is a well-known and expected mechanism. However, the authors also observed a unexpected mechanism that involves the removal of lattice planes at the grain boundaries. Their observations show the importance of using very-high-resolution microscopy to understand the role of grain boundaries during deformation. —BG

Science, abm2612, this issue p. 1261

GRAPHENE

The ABCs of ABC graphene

Stacking graphene layers on top of each other in particular configurations can have a profound effect on the electronic properties of the resulting material. Spectroscopic methods can be used to study the band structure, but the addition of top layers to control the carrier density and improve sample properties makes such measurements tricky. Yang *et al*. overcame these challenges, using infrared spectroscopy in dual-gated trilayer graphene stacked in the so-called ABC configuration and encapsulated in hexagonal boron nitride. The researchers mapped out the band structure features created by the stacking and electron correlations and measured the parameters of the many-body model expected to describe this material. The method may be extended to study other related superlattices. —JS

Science, abg3036, this issue p. 1295

DENDRITIC CELLS

The trouble with increased tolerance

Dendritic cells (DCs) are crucial components of immune responses, and their environment often dictates whether they are pro- or anti-inflammatory. Tang *et al.* investigated the impact of the protein Bat3 in DC function in mouse models of experimental autoimmune encephalomyelitis (EAE) and cancer. Using mice with a deletion of Bat3 in DCs, they found that EAE was attenuated but tumor growth was promoted. These effects were linked to a more tolerant phenotype in the DCs with the Bat3 deletion, which led to more anti-inflammatory T cells. Bat3 deletion was also linked to an altered metabolic and unfolded protein response, leading to increased glucocorticoid production by DCs and subsequent T cell suppression. —DAE

Sci. Immunol. **7**, eabm0631 (2022).

CELL THERAPY

Efficient replacement

Cell therapies have the potential to be effective at treating neurodegenerative conditions. However, the need for local delivery and the poor distribution of the transplanted cells hinders development. Shibuya *et al.* developed an efficient microglia replacement approach in rodents using circulation-derived myeloid cells (CDMCs). The cells broadly incorporated in the brain and generated microglia-like cells more efficiently than bone marrow transplantation. In a mouse model of progressive neurodegeneration, CDMC-mediated microglia replacement reduced cell loss and brain inflammation, improved motor behavior, and extended life span. The results suggest that this approach might be therapeutic in multiple neurological conditions. —MM

Sci. Transl. Med. **14**, eabl9945 (2022).

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**

ILLUSTRATION: ROBERT NEUBECKER

CONSERVATION

Transboundary tree rescue

Magnolia grandis, a critically endangered tree species, inhabits broad-leaved mountain forests along the Sino-Vietnamese border. Blair *et al.* modeled how the tree's preferred habitat is likely to shift in response to climate change in the coming decades, finding losses of suitable habitat in current conservation areas and a likely southward shift in distribution into Vietnam. Targeted conservation management, including population enhancement through nursery cultivation, can be deployed to alleviate potential depletion of the remaining tree populations. These findings highlight the importance of cross-border conservation efforts in a changing climate. —AMS

Front. Biogeogr. **14**, e54662 (2022).

METABOLISM

Re-energizing the aged brain

Nicotinamide adenine dinucleotide (NAD) is an important cofactor in numerous metabolic reactions. NAD concentrations decline with age, which may contribute to age-associated conditions such as Parkinson's disease. Preclinical studies show that replenishing NAD by supplementation with nicotinamide riboside (NR), a biosynthetic precursor to NAD, can promote health span and neuroprotection. Brakedal *et al*. performed a randomized, double-blind phase 1 clinical trial of NR supplementation in 30 patients newly diagnosed with Parkinson's disease. They found that NR supplementation was safe and that concentrations of NAD in the brain increased in most patients receiving NR. These patients had signs of altered

STEM WORKFORCE

Women make better mentors

Previous research on mentoring has demonstrated benefits for mentees and mentors who share demographics. Blaney *et al.* further explored the gender aspect of mentoring relationships, specifically in biology, where gender dynamics are more level. Using structural equation and latent growth curve modeling, the role of adviser gender identity alignment between doctoral students and their advisers in predicting key graduate school outcomes was examined. Among many results, adviser gender indirectly predicted desirable outcomes for women as a function of perceived adviser interaction quality. This relationship was also present for men, suggesting that all students may have more positive outcomes when their adviser is a woman. These results suggest that faculty and administrators should reflect on why students might report lower-quality experiences with male advisers. —MMc

J. Res. Sci. Teach. 10.1002/tea.21761 (2022).

cerebral metabolism and mild clinical improvement, although further testing is needed with a larger cohort to confirm any clinical benefit. —GKA

Cell Metab. **34**, 396 (2022).

EXOPLANETS

Pulsar planets are extremely rare

The first exoplanets were detected in 1992, orbiting the pulsar PSR B1257+12. Few other "pulsar planets" have been found in the following three decades, leaving their abundance poorly constrained. Nițu *et al.* performed a systematic search of 800 pulsars that were monitored for years or decades each, looking for periodic shifts in the radio pulse arrival times caused by an orbiting planet. They found six candidate signals, but none was convincing enough to claim a detection. The results show that less than 0.5% of pulsars host terrestrial planets, which are common around normal stars. Such planets must be destroyed or ejected by the supernovae that form pulsars. —KTS

Mon. Not. R. Astron. Soc. 10.1093/mnras/stac593 (2022).

INORGANIC CHEMISTRY

A fair fight between boron and aluminum

Three-coordinate boron and aluminum compounds are widely used as Lewis acids. Predicting their relative reactivities can be a challenge because the outcome often depends on which Lewis basic counterpart is involved. Kurumada and Yamashita report a molecule with a boron site bonded to a comparably situated aluminum site. Using a combination of crystallography, spectroscopy, and computation, the authors conclude that reaction with carbon monoxide starts at boron due to orbital interactions, whereas Coulombic factors guide dimethylsulfoxide to bind first to aluminum. —JSY

J. Am. Chem. Soc. 10.1021/jacs.2c01580 (2022).

GENE REGULATION

Hemoglobin switching

At birth, many organ functions suddenly change. Lungs shift from being fluid filled to air breathing, and hemoglobin switches from a fetal form into a predominantly adult form. This switch relies on the silencing of the fetal form γ-globin gene and up-regulation of the adult form δ-globin and β-globin genes. The benign condition, hereditary persistence of fetal hemoglobin (HPFH), occurs when large deletions downstream of the fetal globin gene increase its expression. HPFH can be coinherited with sickle cell disease and β-thalassemia, which, interestingly, results in less severe symptoms associated with hemoglobinopathies. Topfer *et al*. found that the promoters for fetal and adult form globin genes compete for the same enhancer called the locus control region. If the promoter for adult form globin genes are removed using CRISPR, then the fetal globin gene loops to the shared enhancer and reactivates fetal form globin expression. Recapitulating mutations that cause HPFH may offer translational avenues for treating blood conditions such as sickle cell anemia. —BAP

Blood. 10.1182/blood.2021014205 (2022).

CELL BIOLOGY

Sneaking out the front door

The spectacular success of mRNA-based vaccines in the past 2 years has spurred interest in delivering intact mRNAs into cells, organs, and tissues. Lipid nanoparticle–associated mRNA (LNP-mRNA) delivery also provides a potentially fruitful avenue for a variety of therapeutics. To understand some of the challenges with this method, Paramasivam *et al.* studied the cell biology of LNP-mRNA delivery. The authors compared the endosomal distribution of six LNP-mRNA formulations in primary human adipocytes, fibroblasts, and HeLa cells. Total LNP-mRNA uptake was not enough to indicate effective delivery because different LNPs varied in their endosomal distribution. The most effective LNP-mRNA delivery occurred soon after internalization from early-recycling endosomes. Super-resolution microscopy elucidated the journey of a single LNP-mRNA and captured events of mRNA escape from endosomal recycling tubules. —SMH

J. Cell Biol. **221**, e202110137 (2022).

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

EVOLUTION

Dynamic adaptation in fruit flies

Ecosystems can experience rapid environmental change, but whether populations can continuously adapt to those changes is unknown. Rudman *et al*. observed rapid parallel evolution in 10 *Drosophila melanogaster* populations over 4 months of seasonal change (from summer to autumn) in Pennsylvania (see the Perspective by Hoffmann and Flatt). Combining a field experiment with laboratory common garden experiments, they observed changes in six phenotypes related to reproductive output or stress tolerance underlain by rapid, genome-wide evolution. The direction of trait and genomic shifts changed over months, in accordance with environmental changes. This study demonstrates the potential for rapid, continuous evolution to changing environmental conditions and highlights the importance of collecting data with a high temporal resolution for observing the effects of fluctuating selection. —BEL

Science, abj7484, this issue p. 1246;
see also abo1817, p. 1226

METABOLISM

Warming up

There are various types of fat cells, including brown and beige adipocytes, that generate heat. These cells activate specialized thermogenic metabolic pathways that use macronutrients from blood and dissipate their energy as heat. Thus, these cells are important to burn up excess calories, and their activity is linked to cardiometabolic health. In a Perspective, Wolfrum and Gerhart-Hines discuss the distinct metabolism of thermogenesis and how we might better understand the pathways involved and how to regulate them. Work in this area may unlock the possibility of targeting thermogenesis to tackle metabolic conditions such as obesity and type 2 diabetes. —GKA

Science, abl7108, this issue p. 1229

CANCER

Detecting cancer early

Many types of cancer are detected at an advanced stage, when treatment options are limited and prognosis is poor. Being able to detect cancers early can substantially improve survival, but this approach comes with challenges, including the possibility of overdiagnosis and overtreatment, which can harm people who would not have developed overt malignancy. In a Review, Crosby *et al*. discuss the importance of cancer early detection and the main challenges that need to be overcome to better understand the early events in tumorigenesis that are detectable in screening tests. The results of these tests can then be reliably interpreted to determine whether an individual requires treatment. —GKA

Science, aay9040, this issue p. 1244

NEUROSCIENCE

Calcium release and neuronal plasticity

What is the role of intracellular calcium release from the endoplasmatic reticulum in neuronal signal processing and in the formation of hippocampal place fields? O'Hare *et al*. used single-cell viral delivery techniques, optogenetics, and in vivo calcium imaging to simultaneously record dendritic and somatic activity of area CA1 pyramidal neurons. Increasing intracellular calcium release increased spatial tuning in apical and, to a lesser extent, in basal CA1 pyramidal cell dendrites. This activity in turn changed place cell responses during learning and memory storage. Intracellular calcium release in concert with circuit-level anatomical features thus shapes and promotes somatic feature selectivity in vivo. —PRS

Science, abm1670, this issue p. 1245

CELL DEATH

Copper induces cell death

Cell death is an essential, finely tuned process that is critical for the removal of damaged and superfluous cells. Multiple forms of programmed and nonprogrammed cell death have been identified, including apoptosis, ferroptosis, and necroptosis. Tsvetkov *et al*. investigated whether abnormal copper ion elevations may sensitize cells toward a previously unidentified death pathway (see the Perspective by Kahlson and Dixon). By performing CRISPR/Cas9 screens, several genes were identified that could protect against copper-induced cell killing. Using genetically modified cells and a mouse model of a copper overload disorder, the researchers report that excess copper promotes the aggregation of lipoylated proteins and links mitochondrial metabolism to copper-dependent death. —PNK

Science, abf0529, this issue p. 1254;
see also abo3959, p. 1231

RADIO ASTRONOMY

Polarized repeating fast radio bursts

Fast radio bursts (FRBs) are intense, millisecond flashes of radio emission from extragalactic sources of unknown origin. Most FRBs are seen only once, but others repeat at irregular intervals and therefore can be followed. Feng *et al*. measured the polarization of five repeating FRBs (see the Perspective by Caleb). They found that each source is polarized at high frequencies but becomes depolarized below a threshold frequency that varies between sources. The authors found that all repeating FRBs are 100% polarized at the source, before the radio waves scatter off complex foreground structures such as supernova remnants. These results constrain theories of the repeating FRB emission mechanism. —KTS

Science, abl7759, this issue p. 1266;
see also abo2353, p. 1227

ORGANIC CHEMISTRY

Galbulimima alkaloids built flat first

The bark of tropical *Galbulimima* plants hosts a variety of complex alkaloids under study for their neurological effects. Their dense frameworks of fused saturated rings have posed a particular challenge to chemical synthesis. Landwehr *et al*. report an efficient synthetic route to three of these alkaloids in which flat, aromatic building blocks are first stitched together and then reduced collectively (see the Perspective by Pitre). Key to the approach was a carefully optimized radical coupling of an aryl bromide with a siloxycyclopropane, which was achieved using nickel–photoredox dual catalysis. —JSY

Science, abn8343, this issue p. 1270;
see also abo2398, p. 1234

MOLECULAR BIOLOGY

A histone variant protects the genome

Histone H3.1 and H3.3 variants are present in proliferative cells of multicellular eukaryotes. Although these two variants are nearly identical, histone H3.1 expression and deposition in chromatin occurs only during DNA replication, which suggests a specific function for this variant during this process. Davarinejad *et al*. discovered that the DNA repair protein TSK/TONSL maintains genome stability during replication by interacting with the H3.1 variant (see the Perspective by Berger). Recognition of H3.1 by the TPR domain of TSK/TONSL regulates the activity of the DNA repair protein, and this role for the H3.1 variant is likely conserved from flowering plants to mammals. —DJ

Science, abm5320, this issue p. 1281; see also abo4219, p. 1232

METALLOENZYMES

Getting the environment just right

Particulate methane monooxygenase (pMMO) is a key enzyme in the metabolism of methane by aerobic, methanotrophic bacteria. There is considerable interest in understanding this reaction, both as a fundamental biogeochemical transformation and for potential biotechnological applications, but progress has been hampered by inactivation of this membrane-bound metalloenzyme upon purification. Koo *et al.* reconstituted natively produced pMMO in lipid nanodisks and found that restoration of the bilayer environment also restored activity of the enzyme. A series of cryo–electron microscopy structures revealed previously missing regions of the enzyme, including conserved residues that coordinate a metal ion near the inner pore of the trimeric complex in a site not seen in previous structures. Lipids play an important role in stabilizing various parts of the structure, and the recovery of activity in these samples is consistent with the newly observed metal-binding site being involved in methane oxidation. —MAF

Science, abm3282, this issue p. 1287

PHYSIOLOGY

From nucleus to kidney dysfunction

Acute kidney injury (AKI) has a high mortality rate. Wang *et al.* found that AKI in mice induced the nuclear export of the catalytic subunit of the kinase DNA-PK (DNA-PKcs) to the cytoplasm, where it phosphorylated a protein called Fis1. Inhibition of DNA-PKcs or expression of a Fis1 mutant that could not be phosphorylated reduced kidney dysfunction in multiple mouse models of AKI. Urinary sediments of AKI patients had increased cytoplasmic amounts of DNA-PKcs and greater phosphorylation of Fis1, both of which were correlated with kidney dysfunction. —WW

Sci. Signal. **15**, eabh1121 (2022).

REVIEW SUMMARY

CANCER

Early detection of cancer

David Crosby*, **Sangeeta Bhatia, Kevin M. Brindle, Lisa M. Coussens, Caroline Dive, Mark Emberton, Sadik Esener, Rebecca C. Fitzgerald, Sanjiv S. Gambhir, Peter Kuhn, Timothy R. Rebbeck, Shankar Balasubramanian***

BACKGROUND: When cancer is detected at the earliest stages, treatment is more effective and survival drastically improves. Yet ~50% of cancers are still only detected at an advanced stage. Improved earlier detection of cancer could substantially increase survival rates. Although recent advances in early detection have saved lives, further innovations and development of early cancer detection approaches are needed. The field is evolving rapidly, owing to advances in biological understanding and an increasing pace of technological progress.

ADVANCES: We highlight five challenges facing the field, current work in those areas, and where more research is needed to make early detection a reality. The first challenge is to build a greater understanding of the biology and behavior of early disease. This will help identify ways to distinguish between consequential, aggressive lesions and inconsequential lesions that will not cause harm. Such insight will be crucial to realizing the potential for early detection to inform treatment decisions and improve survival, while minimizing the risk of overtreatment. Alongside studies in human samples, better models of disease are enabling identification of early signals of tumorigenesis and clarifying the contributions of the immune system and microenvironment to tumor development.

The second challenge is determining the risk of developing cancer. How can we use germline genomic susceptibility, family history, exposures, demographic, and behavioral data to build nuanced risk models to identify who should be tested for cancer and how test results should be interpreted and followed up? Progress is being made to address this challenge through improved understanding of the genomics of cancer risk, integration of that insight with other risk factors, and the development of large-scale population cohorts where risk models can be developed and validated.

The third challenge is finding and validating biomarkers of early cancer. There is considerable difficulty in finding accurate signals of early cancer (which usually exist in very small amounts) amid the noise of normal human physiology. Although progress has historically been slow, many promising early detection markers are emerging, including circulating tumor DNA, circulating tumor cells, proteins, exosomes, and cancer metabolites. Advances in data analysis methodologies (such as machine learning) and integration across marker types in multimodal tests are also accelerating progress.

The fourth challenge is technological. It involves both the iterative improvement of existing approaches and the development of disruptive detection technologies that can very sensitively and specifically identify early biological changes, whether in tissue structure, biochemistry, or function. Powerful molecular analytical technologies and advanced imaging and histopathological methods are increasing the ability to sensitively find earlier tumors, while the use of synthetic markers may help to amplify their signal.

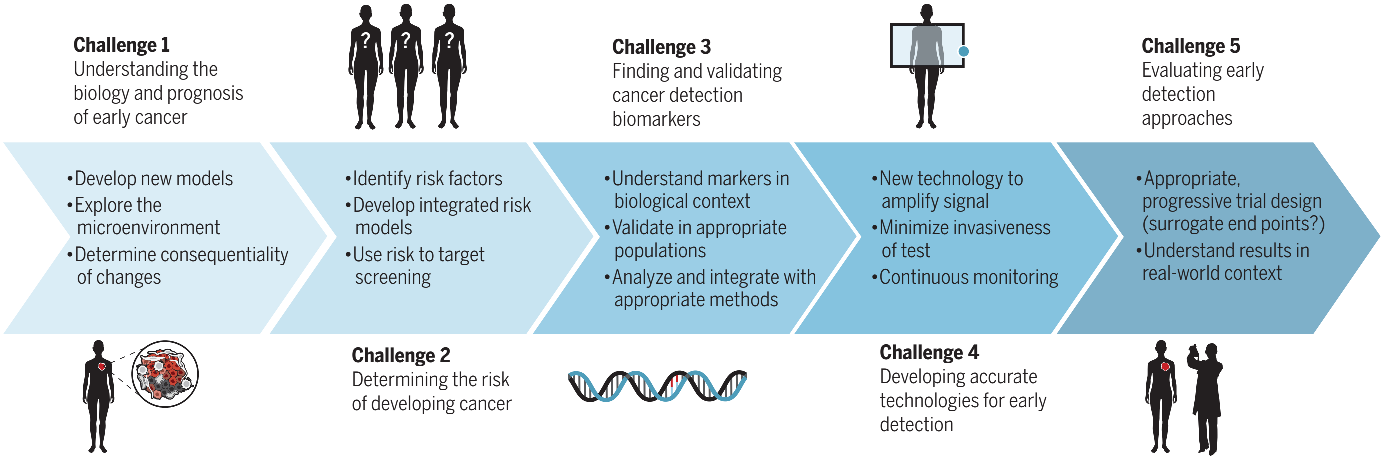
The fifth challenge is how to appropriately evaluate early detection approaches. Translation of biological insights into new diagnostic technologies and execution of clinical trials to validate those advances require substantial time and money. We discuss ways in which that process might be improved.

OUTLOOK: For early detection to deliver transformative progress in cancer survival, wider skill sets beyond cancer biology are essential, including engineers, chemists, physicists, technology developers, and behavioral and computer scientists. Integrated, interdisciplinary collaboration is key to bringing new ideas to address the challenges of early cancer detection. We believe that early detection of cancer is approaching a tipping point, as biological insight and technological capacity are increasing at an unprecedented rate and as public and private funders of research are increasingly willing to invest. This Review discusses the current state of the field and suggests constructive ways forward that build on current progress to deliver effective earlier detection of cancer and appropriate intervention. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: david.crosby@cancer.org.uk (D.C.); sb10031@cam.ac.uk (S.Ba.)

Cite this article as D. Crosby *et al*., *Science* **375**, eaay9040 (2022). DOI: 10.1126/science.aay9040

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.aay9040



The early detection of cancer—challenges and ways forward. This figure summarizes challenges that impede the early detection of cancer and the areas of current research that are helping to overcome them.

REVIEW

CANCER

Early detection of cancer

David Crosby[1]*, Sangeeta Bhatia[2,3], Kevin M. Brindle[4,5], Lisa M. Coussens[6,7], Caroline Dive[8,9], Mark Emberton[10], Sadik Esener[7,11,12], Rebecca C. Fitzgerald[13], Sanjiv S. Gambhir[14]†, Peter Kuhn[15], Timothy R. Rebbeck[16,17], Shankar Balasubramanian[4,18]*

Survival improves when cancer is detected early. However, ~50% of cancers are at an advanced stage when diagnosed. Early detection of cancer or precancerous change allows early intervention to try to slow or prevent cancer development and lethality. To achieve early detection of all cancers, numerous challenges must be overcome. It is vital to better understand who is at greatest risk of developing cancer. We also need to elucidate the biology and trajectory of precancer and early cancer to identify consequential disease that requires intervention. Insights must be translated into sensitive and specific early detection technologies and be appropriately evaluated to support practical clinical implementation. Interdisciplinary collaboration is key; advances in technology and biological understanding highlight that it is time to accelerate early detection research and transform cancer survival.

Cancer is a major global public health problem; there were 10 million deaths from cancer worldwide in 2020 (*1*). It is the second leading cause of death globally, causing one in six deaths (*2*). For nearly all cancers, the chances of survival increase significantly if the disease is detected, diagnosed, and treated at an early stage (*3*) (Fig. 1).

Early detection aims to identify consequential cancer or precancerous change at the earliest time point at which intervention could improve survival or reduce morbidity. Consequential disease will cause mortality or substantial morbidity within the individual's expected remaining life span. Early detection can take place across several windows during the transition from normal cellular activity to dysregulation to cancer; this includes not only detecting cancer itself at an earlier point in its development but also detecting precursor changes (Fig. 2). Screening, which proactively tests asymptomatic people, constitutes a subset of early detection measures. Many of the principles of early detection interact with other points in cancer care, such as detection of minimal residual disease or disease recurrence (Fig. 2). This Review focuses on early detection of primary cancers and precancerous changes in the context of both screening and symptomatic detection.

Early cancer detection research and development have produced tremendous health benefits, for example, through established screening approaches for cervical, breast, and colorectal cancers, which are now diagnosed less frequently at later stages than cancers without established screening (*4*) (Fig. 1). But many cancers, such as esophageal, pancreatic, and ovarian cancers, are still often diagnosed at advanced stages, when prognosis is extremely poor.

Although early detection confers survival advantages in all populations, ~70% of cancer deaths occur in low- and middle-income countries (*2*), often with late diagnosis. For example, the rate of late-stage breast cancer diagnosis in Black sub-Saharan African women remained well above 60% from the 1970s to 2011, whereas in the US, that rate of late diagnosis decreased from ~60% to 32% in Black women over the same period (*5*). Some cancers that have effective early detection tests, such as cervical cancer, have much higher mortality rates in low human development index (HDI) countries compared with high HDI countries (19.8 versus 3.1 deaths per 100,000, respectively), whereas other cancers without effective early detection tests differ less (e.g., stomach cancer, 5.0 versus 4.0 deaths per 100,000, respectively). Late-stage detection of cancer is a global problem that is exacerbated in resource-poor settings, demonstrating that equity is a

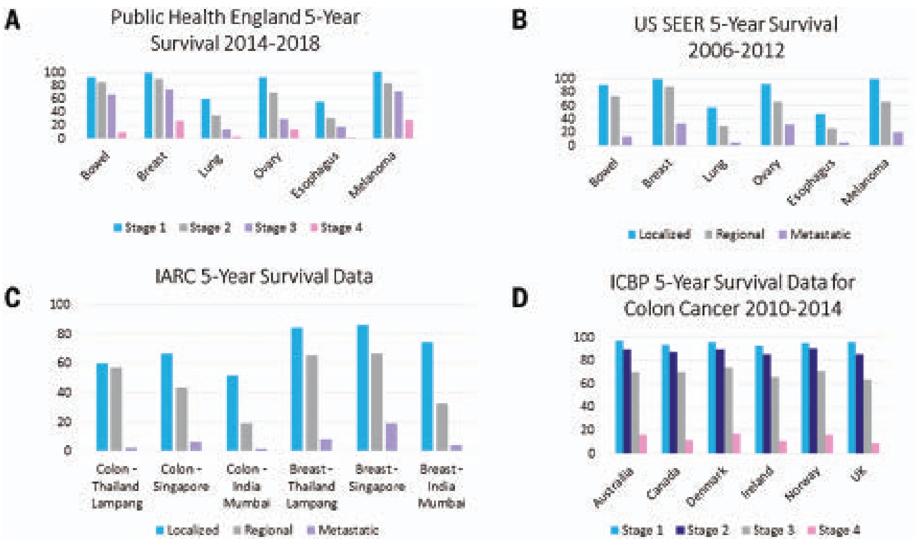


Fig. 1. Patients survive longer when cancer is detected at an early stage. Five-year survival data for bowel, breast, lung, ovarian, and esophageal cancers and melanoma by stage of diagnosis from (**A**) Public Health England (*140*) and (**B**) the US Surveillance, Epidemiology, and End Results (SEER) database (https://seer.cancer.gov). (**C**) Data from the International Agency for Research on Cancer (IARC; https://survcan.iarc.fr/indexsurvcan1.php) shows 5-year survival by stage of diagnosis for colon and breast cancers in Asian countries. (**D**) International comparison [International Cancer Benchmarking Partnership (ICBP) data; https://gco.iarc.fr/survival/survmark] across countries for 5-year survival of colon cancer shows similar trends in percentages of patients surviving early-stage disease versus late-stage disease.

---

[1]Cancer Research UK (CRUK), London, UK. [2]Marble Center for Cancer Nanomedicine, Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA, USA. [4]Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK. [5]Department of Biochemistry, University of Cambridge, Cambridge, UK. [6]Cell, Developmental and Cancer Biology, Oregon Health and Science University, Portland, OR, USA. [7]Knight Cancer Institute, Oregon Health and Science University, Portland, OR, USA. [8]Cancer Research UK Lung Cancer Centre of Excellence at the University of Manchester and University College London, University of Manchester, Manchester, UK. [9]CRUK Manchester Institute Cancer Biomarker Centre, University of Manchester, Manchester, UK. [10]Division of Surgery and Interventional Science, University College London, London, UK. [11]Department of Biomedical Engineering, School of Medicine, Oregon Health and Science University, Portland, OR, USA. [12]Cancer Early Detection Advanced Research Center, Oregon Health and Science University, Portland, OR, USA. [13]Medical Research Council (MRC) Cancer Unit, Hutchison/MRC Research Centre, University of Cambridge, Cambridge, UK. [14]Department of Radiology, Molecular Imaging Program at Stanford, Stanford University, Stanford, CA, USA. [15]USC Michelson Center Convergent Science Institute in Cancer, University of Southern California, Los Angeles, CA, USA. [16]Division of Population Science, Dana-Farber Cancer Institute, Boston, MA, USA. [17]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [18]Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge, UK.
***Corresponding author. Email: david.crosby@cancer.org.uk (D.C.); sb10031@cam.ac.uk (S.Ba.)**
†Deceased.

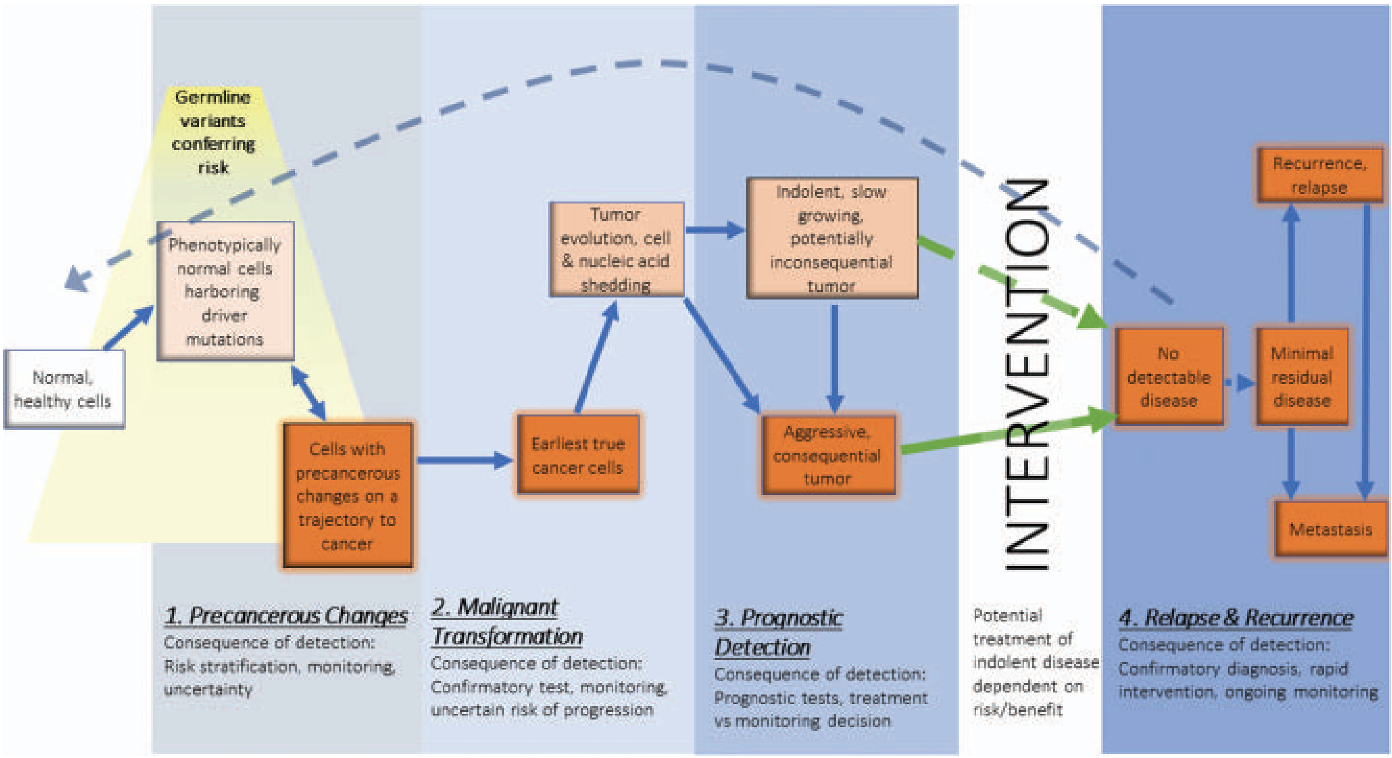


Fig. 2. Windows for early detection across the course of cancer progression. Cancer evolves through various stages, offering multiple windows for early detection. Detection at each stage presents different information and choices, with the consequences of detection dependent on the level of information provided by the subsequent test(s) and the level of certainty around whether the disease will be consequential.

considerable challenge (*6*–*8*). Patients diagnosed with later-stage cancer can miss the window for curative intervention, and expensive later-stage systemic treatments are often associated with severe side effects and worse outcomes (Fig. 1). Further research to build on early detection successes and extend these into other cancer types could transform patient outcomes.

The challenges facing early detection research fall into five broad categories. First, understanding the biology of early cancer: What should we look for, and, once found, how can we know which early lesions will progress to become aggressive, consequential disease versus indolent, inconsequential disease? Second, determining risk: There are substantial challenges in knowing which populations or individuals are at greater risk of developing cancer and therefore in deciding who should be tested and how tests should be interpreted and acted on. Third, finding and validating biomarkers: Early tumors are miniscule—discovering sensitive markers of their presence and robustly validating them presents an archetypal needle-in-a-haystack challenge. Fourth, developing accurate technologies: there is a considerable challenge in developing technologies that are sensitive enough to detect markers of early cancers and specific enough to avoid false alarms or overtreatment for inconsequential disease. Fifth, evaluating early detection approaches appropriately: The ultimate challenge is to robustly demonstrate that a new early detection approach can indeed detect cancers early and ultimately save lives. The relative scarcity of cancer in the general population can make this a difficult, prolonged, and extremely expensive process.

There is a general need for accurate early detection technologies that address the issues of cost, access and scaling, public acceptance of testing, and integration of diagnostics with public health infrastructure and decision-making. The point-of-care tools and privacy-compliant telehealth solutions that have emerged to meet the COVID-19 pandemic crisis may also help advance the implementation of early cancer detection. Early detection approaches must address, rather than exacerbate, health inequities and must achieve a positive balance of benefit to harm (through overdiagnosis, unnecessary invasive follow-up, and overtreatment for inconsequential disease). In this Review, we describe the diverse research challenges and propose ways of achieving early detection of cancer.

Challenge 1: Understanding the biology of early cancer

There is a continuum in tumorigenesis from normal to dysregulated to cancerous. A key challenge is to understand this biology so that we can predict the future trajectory of the changes we detect and determine when early disease becomes consequential and/or lethal.

The cancer continuum and transition to lethality

Cancer evolves from early inconsequential dysregulation in molecular and cellular phenotypes, to malignant transformation where critical changes in a cell's genome or epigenome culminate in a hallmark series of abnormal features that define cancer, to potentially lethal invasion and metastasis and ongoing cellular evolution and diversification (*9*). Windows of opportunity, as well as challenges, for cancer detection exist across this continuum (Fig. 2). The transition rate through these stages depends on the cancer type, therefore, understanding this timeline can help pinpoint the optimal time for detection and intervention.

Annual screening may not detect fast, aggressive cancers that develop between screening visits (*10*). Conversely, slow-growing cancers undergoing malignant transition over several years can be tracked with active surveillance and screening of at-risk populations. Some cancers follow a clear path from precursor condition to malignancy, such as polyps preceding colon cancer. However, not all precursors will progress to cancer, and not all cancers will be consequential. For example, the precancerous condition monoclonal gammopathy of undetermined significance (MGUS) has an average risk of developing into multiple myeloma (a lethal cancer) of only 1% per year (*11*), and the risk of Barrett's esophagus developing into esophageal cancer is ~0.3% per year (*12*). We do not fully understand which lesions will progress to consequential disease and which will not.

What confers lethality and its timing? A cancer can theoretically be traced, by means of a phylogenetic tree, back to a single cell. This single cell arises from a specific set of conditions, including the tissue microenvironment and the immune system. Each organ system presents a different environment, with some mutations causing a potentially lethal tumor in one organ context but not another (*13*). The picture also changes within individuals as a result of aging. A tumor-permissive environment can be created by cellular and molecular changes in noncancerous cells during aging, such as biophysical alterations in the extracellular matrix, changes in secreted factors, and changes in the immune system (*14*). What are the transitions leading to that initial cancer cell and then the changes that engender a consequential tumor, both within the cell and its interactions with its microenvironment? The early evolution of most cancers cannot easily be observed in people owing to clinical presentation at advanced disease stages and tissue sampling difficulties when monitoring precancers. Blood cancers are an exception, where the ease of blood sampling has allowed better understanding of clonal hematopoiesis (*15*). For example, all multiple myelomas will have progressed from MGUS. Chromosomal and other mutational changes can be monitored in MGUS patients and may highlight patients who are progressing from MGUS to smoldering myeloma to malignant multiple myeloma (*16*). Clinical trials suggest that patients who undergo early detection of MGUS progression may benefit from therapeutic intervention at the stage of smoldering myeloma rather than waiting for symptomatic malignant myeloma with end-organ damage (*17*). This demonstrates how detection and molecular stratification of a preneoplastic lesion (Fig. 2) can trigger intervention before clinically observed definite malignancy. Given that not all patients presenting with multiple myeloma will have a prior clinical diagnosis of MGUS, monitoring changes in MGUS patients will not catch every case, although it does give a paradigm to study and exploit the biology of transition from precancer to cancer.

The transition from a normal state to cancer is also affected by a cell's microenvironment. The microenvironment includes host immune cells, mesenchymal support cells, vascular cells, extracellular matrix, and secreted proteins and exists in various states of hypoxia and pH. The microenvironment surrounding a would-be tumor cell can contribute to tumor progression, determining whether that cell remains localized or spreads aggressively. An early tumor may also induce detectable changes in its microenvironment, generating potential biomarkers for detection.

The immune system is a crucial regulator and indicator of the initiation and progression of early tumors (*18*). For example, the spatial positioning of tumor-infiltrating leukocytes with regard to the tumor can, in some cancer types, indicate how invasive a tumor is (*19*), and leukocyte-based biomarkers may be used to identify residual disease after therapy or to predict response to therapies (*20*). However, it is becoming clear that immune cells or their products may themselves be useful for early detection (*21*). As discussed in the Challenge 3 section below, the very small size of the earliest tumors means that any biomarkers they shed into the circulation will exist in very small amounts, impeding detection. The human immune system could act as a signal amplifier (each tumor cell being potentially exposed to many immune cells). This exquisitely sensitive apparatus could be harnessed to signal the presence of a cancer. Immune system biomarkers currently under investigation as early cancer detectors include the overall immune contexture in the peripheral circulation, autoantibodies (*22*), T cell repertoires (*23*), and leukocyte-shed exosomes.

Biological models of disease

Because we cannot easily observe the first tumor cell to emerge in humans, cancer models have been developed to probe the mechanisms underlying tumor initiation (*24*–*27*). However, there are few models of very early cancer or premalignant disease that faithfully reproduce somatic events leading to disease in immunocompetent native tissue microenvironments.

First-generation transgenic models of human cancer progression (*28*) afforded initial glimpses of tissue- and organ-specific biology of neoplastic progression. Although such studies have revealed tumor cell–intrinsic (*29*, *30*) and tumor cell–extrinsic characteristics (*9*, *31*) that support malignancy, these models have substantial drawbacks, such as rapid progression, and phenotypes that are frequently fully penetrant. Therefore, these models do not accurately recapitulate human disease.

Models of early cancer have been improved by developing immunocompetent mouse models with constitutive and conditional mutations in multiple cancer-associated genes, as well as embracing tumor microenvironment and epigenetic regulators. For example, mouse models allowing exploration of early tumorigenesis and that more closely recapitulate human disease (including immunocompetent and conditional expression models) now exist for nonmelanoma squamous cell carcinoma (*32*), pancreatic adenocarcinoma (*33*), colon cancer (*34*), and lung adenocarcinoma (*35*, *36*). Further insights will come from the next wave of model systems using approaches including circulating tumor cell patient-derived explants (*37*, *38*), patient-derived xenografts, and creation of complex organoids involving multiple cell types (*39*). However, many patient-derived xenograft models use samples from advanced human disease implanted into immunocompromised mice, which may not reflect truly early disease processes or the important role of the immune response to early lesions. With increasing sophistication, the interplay of patient-derived models and advanced nonhuman model systems can provide a path to greater understanding of early cancer biology, early detection markers, prognosis, and appropriate interventions for early cancers.

Challenge 2: Determining risk of developing cancer

Understanding whom, how, and when to test and also how test results should be interpreted requires understanding of individual cancer risk. Early detection strategies will not be of equal value to everyone. Therefore, it is important to identify the people at elevated risk of cancer and to tailor an early detection strategy to that group to maximize the benefits of early detection and minimize the harms (through over- or underdiagnosis and treatment) (*40*).

Risk models

Risk assessment models can identify individuals or populations at increased risk for a specific cancer or cancers. Risk stratification includes information about age, familial history, exposures, and lifestyle (*41*), which can be augmented by genetic screening to detect variants in genes associated with cancer. This strategy is exemplified by breast cancer risk prediction models used to stratify women into higher-risk categories and toward genetic testing for inherited cancer susceptibility (*42*, *43*). Women with inherited *BRCA1* or *BRCA2* pathogenic variants associated with increased risk of breast and ovarian cancers are candidates for chemoprevention with selective estrogen receptor modifiers, risk-reducing surgery, or enhanced breast magnetic resonance imaging (MRI) screening to enable earlier detection. Currently, very few high-risk single genes (such as *BRCA1* and *BRCA2*) trigger such action, however polygenic risk scores are being explored, which consider the risk conferred by multiple genetic variants (*44*). Better precision is needed in the identification of high-risk people who require screening for early cancer detection. Improved accuracy of risk modeling will be enabled by the discovery and use of more-informative genomic and phenotypic (e.g., breast density) markers of risk, integrated into multifactorial models that also consider, for example, family history and behavioral factors. It is crucial that risk models are evaluated by experts using the appropriate statistical methodologies and validated in independent datasets (*45*).

Constructing improved risk stratification models requires data and biological samples from large cohorts, ideally in prediagnostic populations that are followed for any cancer diagnoses. Current examples include the UK Our Future Health initiative, which will follow 5 million volunteers (https://ourfuturehealth.org.uk/); Project Baseline in the US, following 10,000 volunteers (www.projectbaseline.com); the Asia Cohort Consortium, following at least 1 million volunteers (www.asiacohort.org); and the European Prospective Investigation into Cancer and Nutrition (EPIC) study, following >500,000 volunteers (https://epic.iarc.fr/). These longitudinal studies in healthy volunteers could help understand the hidden variability between healthy individuals and discover, validate, and contextualize early disease signals. Ultimately, these studies could identify factors to stratify healthy individuals into groups at risk of developing certain cancers.

Screening at-risk populations

Once validated risk models have identified the at-risk populations, these individuals can be invited to participate in screening programs, where available. Screening aims to detect early cancer by inviting asymptomatic, ostensibly healthy people for testing. Ideally, cancer screening should be minimally invasive or noninvasive, low cost, and provide minimal false negatives or positives to minimize harm and maximize benefits of screening. Several existing screening tests improve cancer-specific mortality or overall mortality, including mammography for breast cancer (*46*), the Pap smear for cervical cancer (*47*), colonoscopy for colorectal cancer (*48*), and low-dose computed tomography (CT) (*49*) for lung cancer. Although effective, these technologies are not necessarily minimally invasive, low cost, or highly sensitive and specific. Nor do these tests reach all the at-risk populations concerned. For example, in the US as of 2019, <5% of eligible individuals have been screened for lung cancer (*50*), owing to incomplete implementation of screening in health care systems and low individual compliance.

For screening to be successful, the follow-up diagnostic workup must be feasible and risk-appropriate. For example, breast nodule biopsy (triggered by a positive mammogram) is a low-risk outpatient procedure. Conversely, lung biopsy (triggered by a positive lung CT screen) is highly invasive and relatively high risk. The performance characteristics of the primary screening test, and the threshold set for a positive or negative test result, must be calibrated against the consequences of a positive result. Therefore, any early detection strategy should give rise to actionable, evidence-based follow-up.

Challenge 3: Finding and validating cancer detection biomarkers

A key challenge is how to detect the very small signal of the earliest cancers amid the noise of normal human biology. Two fundamental measures of a diagnostic test are sensitivity and specificity. Sensitivity is the ability of a test to correctly identify those with the condition being tested for (the true positive rate); a test with higher sensitivity will miss fewer cases (i.e., there will be fewer false negatives). Specificity is the ability of a test to correctly identify those individuals without the condition (the true negative rate); a test with high specificity does not give a positive result when the condition is not present (i.e., does not give false positives) (*51*). Sensitivity and specificity depend on both the technology used in the test and the biomarker(s) being measured. Two other key measures are positive predictive value (PPV), which is the probability that individuals who test positive actually have the condition, and negative predictive value, which is the probability that individuals who test negative do not have the condition (*52*). The target values of these parameters will depend on the intended circumstance of use of the test and on the prevalence of the particular cancer being tested for in a given population.

Challenges in biomarker validation

Many biomarkers for early cancer detection have been proposed, but few have been validated in large trials. For example, elevated prostate-specific antigen (PSA) in the blood was a candidate prostate cancer early detection biomarker. However, PSA varies greatly between individuals and within individuals as they age (or as they develop other nonmalignant prostate conditions), leading to the potential for overdiagnosis, unnecessary diagnostic workup (including invasive biopsy, which confers risk), and overtreatment of inconsequential disease (which incurs potential adverse effects without increasing survival) (*53*, *54*). As such, PSA is not generally recommended as a primary, population-level screen (Fig. 3). Another example of a blood marker for cancer that showed promise was CA-125 for ovarian cancer; while use of this marker increased the number of early-stage diagnoses and decreased the number of late-stage diagnoses, this was not accompanied by reduction in mortality (*55*).

Even if validated, highly specific biomarkers can display dichotomy when taken out of context. For example, in colorectal cancer, KRAS mutations are strongly associated with disease progression (*56*), but in the pancreas, many neoplasms carrying KRAS mutations are not malignant (*57*). A useful biomarker must provide enough prognostic, actionable information to inform clinical decision-making.

Promising biomarkers

Biomarkers of early cancer include visible structural changes to tissue and biochemical changes. Minimally invasive sampling methods are preferred, especially where repeated samples from healthy and at-risk individuals are required. In practice, this includes imaging; sampling body fluids such as blood, saliva, or urine (*58*); and sampling tissues via swabs or brushings. Exhaled breath is another source of biomarkers, specifically volatile organic compound (VOC) signatures of cancer and associated metabolites (*59*).

Liquid biopsies (sampling of body fluids) can be used to identify a wide range of substances indicative of cancer, derived either from the tumor itself or from the body's response to the tumor. For example, nucleic acid fragments called cell-free DNA (cfDNA) enter the blood during cellular apoptosis or necrosis. In cancer patients, a portion of the cfDNA, called circulating tumor DNA (ctDNA), is derived from the tumor. Analysis of ctDNA has shown promise for personalized mutation profiling and monitoring of patients with advanced cancers (*58*, *60*), in whom ctDNA levels are relatively high. A key challenge is that ctDNA and indeed all biochemical cancer biomarkers are present at extremely low concentrations in early-stage cancer. New approaches are needed to improve on current limits of detection to address this limitation.

Human genome sequencing (*61*, *62*) has provided unprecedented insights into cancer genomes (*63*, *64*). Although mostly focused on advanced cancer, these studies have elucidated patterns of genetic variants across cancers, some of which may also be present in early tumors. These patterns can provide a basis for detection, stratification, and treatment of cancers (*65*). Liquid biopsy tests based on cancer-associated mutations in ctDNA are showing promise in early detection (*66*). However, it is increasingly clear that phenotypically normal tissue also harbors a range of somatic mutations that might normally be considered indicative of cancer or to be drivers of cancer genesis (*67*). Researchers developing early detection approaches must be mindful of this—how can we define what a normal background of mutations is, as distinct from a consequential cancer signal?

Epigenetic modifications of DNA provide another source of early detection biomarkers. These include cancer-specific DNA methylation profiles (*68*), noncoding RNAs (*69*), small regulatory RNAs, and the DNA modification 5-hydroxymethylcytosine (*70*). One promising approach analyzes methylation patterns of cfDNA in blood (*71*) and is now entering large-scale prospective clinical trials in the UK (NCT03934866) and the US (NCT04241796). Another emerging technique is based on the observation that fragmentation patterns in

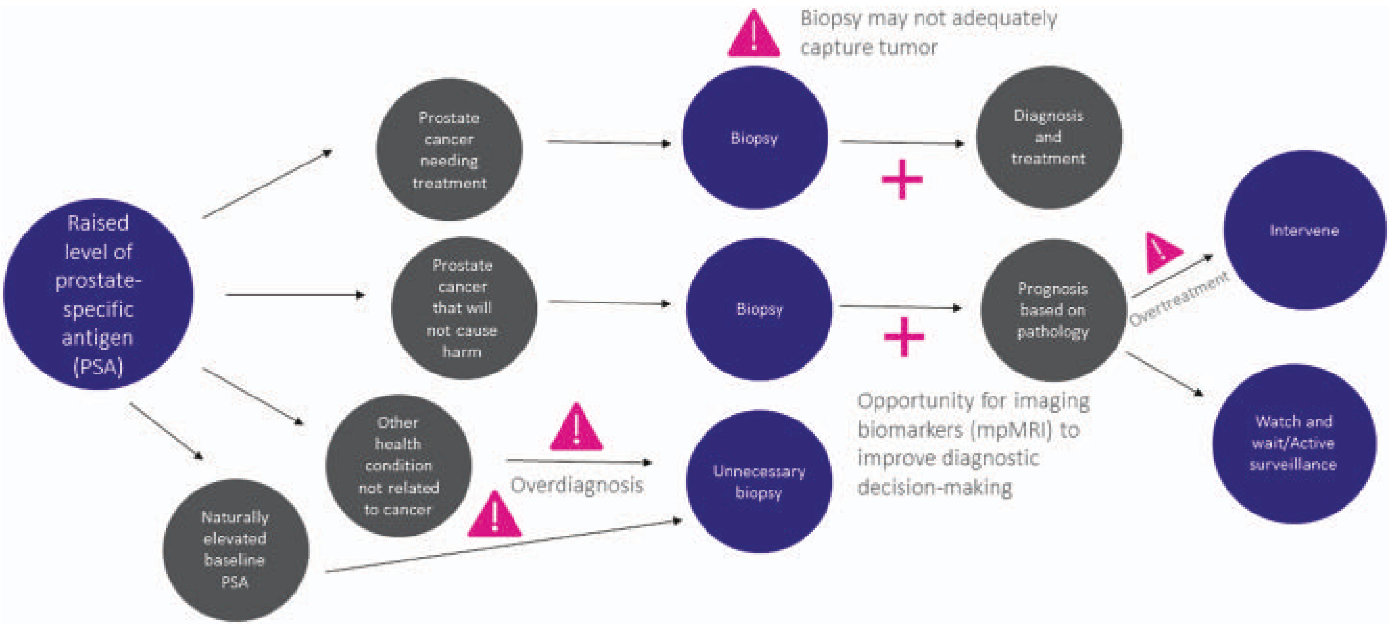


Fig. 3. Prostate cancer detection is a cautionary tale for overdiagnosis and overtreatment. There are various consequences of PSA testing when it is used as a screening tool. An elevated PSA level is not considered useful for prostate cancer screening owing to false positives and detection of inconsequential cancers that will not cause harm in an individual's lifetime. Subsequent biopsy is also imperfect because it does not always capture the tumor and may not distinguish indolent from aggressive cancers. The introduction of imaging biomarkers [e.g., multiparametric magnetic resonance imaging (mpMRI)] combined with pathology (at the junctures indicated by the red plus signs in the figure) has improved prognosis through better stratification of disease.

cfDNA differ between people with and without cancer, and between different cancer types (*72*).

Other potential detection biomarkers include circulating tumor cells (*73*), exosomes (*74*), cell fusions (*75*), metabolites (*76*), and proteins (*77*). These complement DNA sequencing for the discovery and exploitation of cancer-specific signatures (*78*–*80*). Furthermore, certain microbes may confer susceptibility to certain cancers (*81*–*84*), yielding another potential pool of biomarkers. Various types of signal modalities are in clinical use or under development for early cancer detection (Fig. 4). It is possible that multimodal testing will ultimately achieve higher sensitivity and specificity for early cancer than a test that uses a single type of biomarker.

Multimodal testing can be sequential or parallel. Sequential testing cascades from tests indicating risk to confirmatory test(s) of another modality. Although effective (e.g., colorectal cancer screening) (Fig. 5), this results in long, complex diagnostic journeys. Parallel testing integrates data from different modalities to provide the diagnostic signal, for example, detection of the same cancer through measurement of ctDNA, metabolomics, and imaging. Parallel testing has improved the accuracy of liquid biopsy tests in blood (*80*), urine (*85*), and cervical swabs (*86*). Another approach has been to profile both ctDNA mutations and serum protein biomarkers (*80*, *87*), with further improvement achieved by also adding positron emission tomography–computed tomography (PET-CT) imaging (*88*). A prominent example of a successful multimodal cancer detection test combines an assay for fecal blood with a test for known cancer-associated DNA mutations, for improved colorectal cancer screening (*89*) compared with the single fecal hemoglobin test, and is now in clinical use.

Data analytic methods

New computational tools will be vital for analyzing, integrating, and using the data generated by diagnostics. Artificial intelligence (AI) and machine learning (ML) approaches, such as support-vector machine and neural network models, can discover cancer biomarkers, detect cancer-specific signatures in high-dimensional datasets, and build prospective statistical classifiers for evaluating diagnostic performance in independent cohorts (*43*). Such approaches offer exciting avenues for progress but are also fraught with potential challenges, of which researchers should be mindful. Many AI and ML models are criticized for being "black box," that is, unable to explain why the features (e.g., biomarkers) have been selected by the model; the creation of fully interpretable models would be advantageous (*90*). AI and ML models are often developed (or trained) on datasets derived from selected populations that do not represent the real population where the AI-derived test would be used, so they cannot be extrapolated to real-life conditions (*91*). Some AI and ML models are of poor design and insufficient sample size, risking bias and overfitting (*92*). The quality of design and reporting of some trials of AI approaches can also be suboptimal, calling into question the validity of their claims. Design aspects, such as not being prospective, being at high risk of bias, lacking appropriate transparency on data and code, lacking adequate comparator groups, and deviating from existing reporting standards can jeopardize reliability (*93*). In some cases, AI and ML methodology might simply not have advantages over statistical methods such as logistic regression (*94*)—the right tools should be used for the intended purpose.

Challenge 4: Developing accurate technologies for early detection

Developing technologies with the sensitivity to detect the earliest tumors and the specificity to minimize false positives is a key challenge. The emergence of new technologies is enabling early cancer detection with increasing accuracy. One early detection goal is to detect emerging solid tumors that are susceptible to therapy and unlikely to have metastasized. This usually means prior to development of tumor microenvironments that support enhanced angiogenesis and before suppression of antitumor immunity (*27*, *31*, *95*), when the tumor is roughly a millimeter in diameter (comprising 10^5 to 10^6 cells). Most imaging technologies in clinical use or development cannot detect such tiny tumors, but new in vivo imaging instruments such as 10.5T MRI (*96*) are continuously pushing the limits.

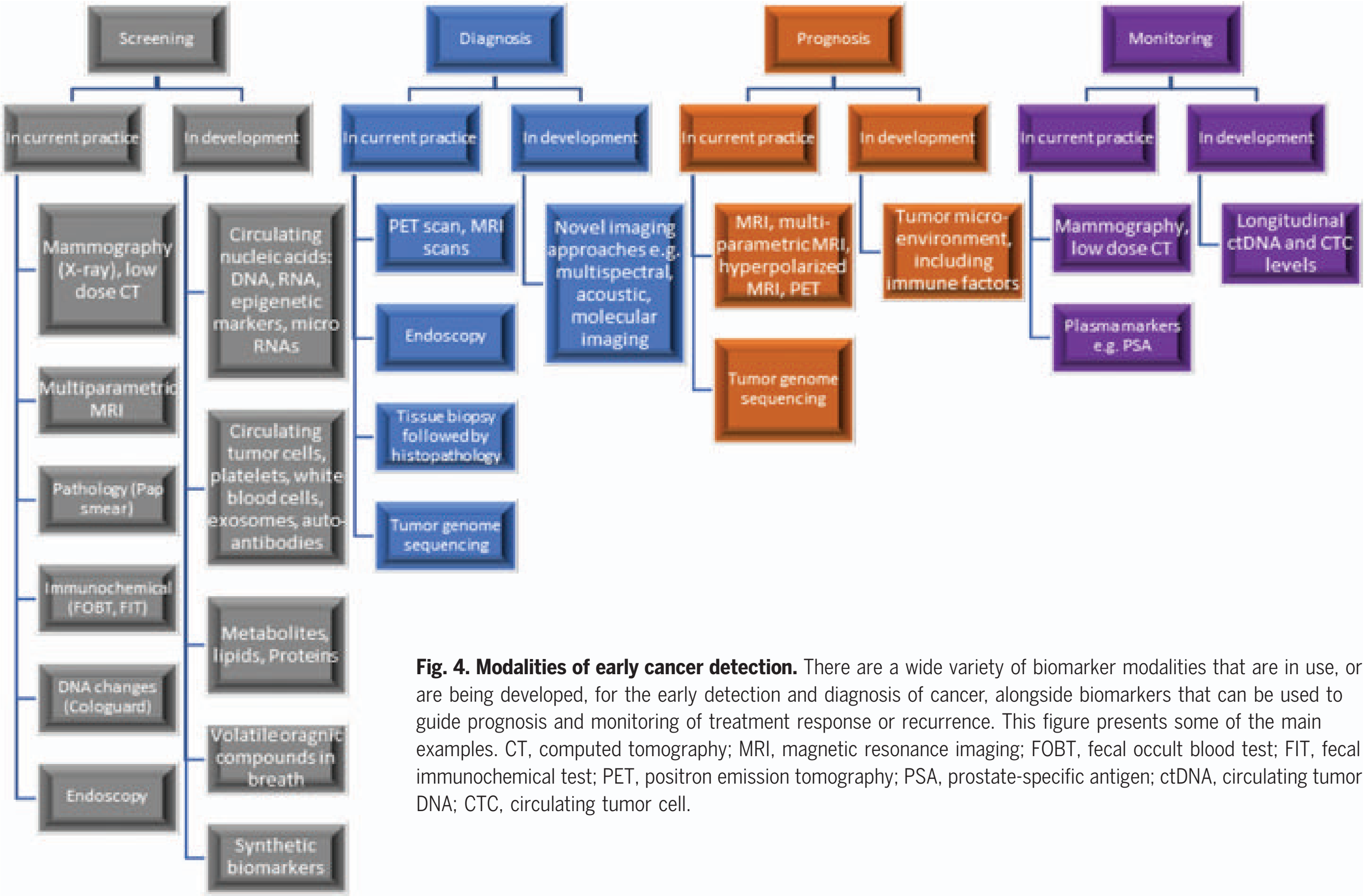


Fig. 4. Modalities of early cancer detection. There are a wide variety of biomarker modalities that are in use, or are being developed, for the early detection and diagnosis of cancer, alongside biomarkers that can be used to guide prognosis and monitoring of treatment response or recurrence. This figure presents some of the main examples. CT, computed tomography; MRI, magnetic resonance imaging; FOBT, fecal occult blood test; FIT, fecal immunochemical test; PET, positron emission tomography; PSA, prostate-specific antigen; ctDNA, circulating tumor DNA; CTC, circulating tumor cell.

New technologies

Sensitivity is being improved by recent technologies that detect tumor metabolites and other secondary products (Fig. 4) that are relatively more abundant than tumor cells. This can be augmented by highly specific probes, such as tumor-specific antibodies or peptides that are radio-labeled to increase signal. Other strategies include engineered diagnostics that are selectively activated in the presence of disease, such as molecular (*85*, *97*) and biological (*98*) sensors that profile the in vivo tumor microenvironment to generate synthetic biomarkers of disease. Activity-based diagnostics use enzyme activity to generate exogenous biomarkers that signal the presence of cancer. For instance, nanoparticles have been developed that are cleaved by dysregulated protease activity in cancer cells to generate urinary reporters (*99*), and cancer-associated enzymes can metabolize exogenous VOC probes to produce volatile reporters for noninvasive detection (*97*). New synthetic biology tools include engineered probiotic (*100*) and immune cell (*101*) diagnostics for tumor detection via amplified, activity-based readouts.

Developments in material engineering and microfabrication have yielded devices that can emulate physiological microenvironments to

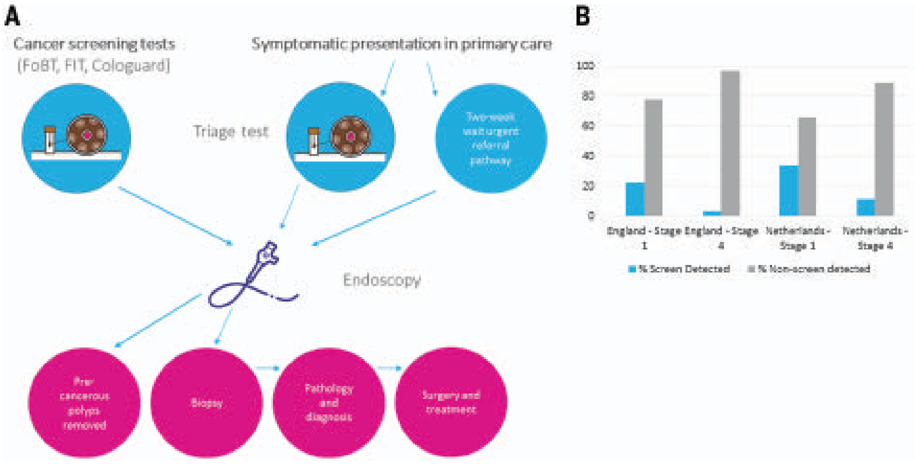


Fig. 5. Colorectal cancer screening is an early detection success story. (**A**) Screening for colorectal cancer (CRC) relies on a cascade of diagnostic tests (fecal screening to endoscopy to biopsy and histopathology) that can lead to (**B**) the detection of cancers at an earlier stage. This screening has transformed CRC into a treatable cancer with increased survival rates when the cancer is caught early (see Fig. 1). Population screening programs have relied on fecal occult blood tests (FOBTs) that measure gastrointestinal bleeding. In the UK, a FOBT was recently replaced by the fecal immunochemical test (FIT—a more accurate method of detecting blood in feces), and in the US, screening also includes the Cologuard FIT-DNA test, which looks for cancer-associated DNA mutations in the feces in addition to the FIT component. Positive results from fecal screening tests usually then cascade to endoscopic examination and, where appropriate, intervention. Most CRC, however, is still diagnosed through presentation to primary care and urgent referral routes, where symptomatic presentation is often associated with a later disease stage.

probe tumor biology and isolate circulating tumor cells (CTCs) and extracellular vesicles from patient samples. Notable examples include label-free capture of CTC clusters (*73*, *102*) and ultrasensitive detection of circulating exosomes with microfluidic chips or external hardware (*103*, *104*). Miniaturization has enabled new sensing approaches using wearables and implantable devices where personalized health data can inform the prevention or interception of certain diseases (*104*). More robust integration of device engineering with downstream molecular profiling technologies will help validate the relevance of these approaches for early detection.

Imaging technologies

Contemporary imaging technologies can only visualize tumors containing more than 10^9 cells; this will miss many of the smaller (i.e., earliest) tumors. Imaging of tissue morphology is currently used in breast cancer screening in the form of x-ray mammography, and low-dose CT is increasingly being used to detect early-stage lung cancer in high-risk groups (*105*). Although these techniques can be used for screening, as they are relatively quick and low cost, they are subject to limited resolution and also confer risk to the patient given their use of ionizing radiation. More advanced imaging modalities are not currently routinely used in primary screening owing to high cost and low availability.

Molecular imaging technologies, such as MRI (*106*) and PET (*107*), can perform early diagnosis and staging. Enhanced variations on these technologies provide the possibility of greater sensitivity, specificity, or PPV, for example, time-of-flight PET (*108*), where transit times of the photons emitted by the object generating the image signal provide a greater signal-to-noise ratio, and hyperpolarized MRI (*109*, *110*), where hyperpolarized carbon-13–containing molecules enable the collection of perfusion and metabolic information in addition to structural images.

Using imaging to examine multiple properties of a lesion can enhance the detection and classification of early lesions. For example, multiparametric MRI of the prostate provides information on prostate volume, cellularity, and vascularity; this can distinguish benign lesions from aggressive tumors requiring intervention (*111*). Imaging has the advantage of being noninvasive and easily repeatable to detect growing tumors. For example, lung cancer screening with low-dose CT repeated over time can distinguish benign lung nodules of low malignant potential from early lung cancer nodules (*105*).

Computer-assisted diagnostic systems help radiologists to interpret images (*112*). Computer-driven feature extraction can exploit differences in texture and shape that the naked eye cannot see. Digital attributes of the suspect lesion are called radiomic features and may contain indirect information about the underlying histopathology (*113*). This is an area where there is an opportunity for AI and ML to help detect cancer (*114*) and to predict risk of progression (*115*), although issues of transparency and reproducibility must be addressed (*116*). The application of AI in imaging will require large volumes of well-annotated image data, acquired under standardized conditions, representing all populations equitably, and made widely available by means of curated image repositories.

Photoacoustic imaging exposes the region of interest to pulsed laser light of a given wavelength, generating a sound that is measured by microphones or piezoelectric sensors. The level of detail and resolution of the tissue is higher than that of all other types of imaging and is free of ionizing radiation. The challenge is depth of penetration and miniaturization for clinical use (*117*). Visible light imaging through endoscopy has been a mainstay of early detection (e.g., in the colon and lung). The emerging fluorescence endoscopy technique, along with a fluorescent molecular imaging probe, has been used for enhanced detection of lesions in patients with Barrett's esophagus (*118*) and of neoplastic polyps in the colon (*119*).

Histopathology and AI

After initial detection by biomarkers and/or imaging, histopathology is a key confirmatory diagnostic and prognostic stage of the early detection paradigm. The application of ML techniques to digitized slides can increase sensitivity; reduce subjectivity and inter-reader variation; and predict prognosis, recurrence, and tumor susceptibility to treatment (*120*). In some cases, such as Barrett's esophagus dysplasia, bowel polyps, and cervical neoplasia, pathologists examine a precancerous condition with the aim of identifying the transition to early cancer. Digital pathology and AI could help improve test turnaround times and diagnostic accuracy, detecting early signs of cancer and providing data for further research (*121*). Current challenges in digital pathology include handling artifacts, overcoming sample variability, lack of binary variables where a diagnosis may require a risk score, and combining samples across multiple sites and cohorts.

Challenge 5: Evaluating early detection approaches

There are many challenges around the design and methodology of trials of early detection approaches. Trials must be carefully designed to address the relevant population and measure the appropriate end points to provide statistically robust evidence to change practice. Early detection trials differ from the better-known clinical trials for therapeutics and require specialist statistical expertise to inform study design and appropriately powered sample size. For example, the statistical power of early detection trials is affected by factors that do not exist in therapeutic trials, such as the number of times an individual is tested, the time between tests, and the ages at which testing will be performed (*122*). The end points to be considered in diagnostic trials differ from those in therapeutic trials, as do regulatory approval pathways. However, the main challenges to the delivery of early detection trials lie in their scale and interpretation.

The scale of early detection trials

Currently, regulatory or reimbursement decisions on the adoption of cancer screening tests are generally based on impact on mortality: Does the use of the screening test mean fewer deaths from cancer than in an unscreened population? Demonstrating this requires very large numbers of participants (given the comparatively low incidence rate of cancers in an asymptomatic population) and very long timelines (given the potential lag between commencement of the trial, a given individual developing cancer, and that cancer resulting in death). For example, the trials assessing low-dose CT screening for lung cancer in heavy smokers took 7 years and 53,454 participants in the US (*123*) and more than 10 years with 15,789 participants in Europe (*124*). In a more general population (lacking the greatly increased cancer risk of heavy smoking), even greater numbers of participants are needed. For example, trials assessing screening for ovarian and prostate cancers involved more than 200,000 women (*125*) and 184,000 men, respectively (*126*). This scale makes most early detection trials multicenter by default. Clinical trial networks such as those sponsored by the European Organisation for Research and Treatment of Cancer (EORTC) and the US-based National Cancer Institute (NCI) can help to facilitate and accelerate such large trials.

Another attractive option is embedding research into screening programs, taking advantage of existing screening infrastructure. This can, for example, be done using the stepped-wedge design, where observations are initially collected during a baseline period in which no participants are exposed to the intervention (i.e., the new screening test under investigation). After this baseline period, at regular intervals (or steps) participants (or groups of participants) are randomized to receive the intervention; these ascending steps continue until all participants have received the intervention (*127*).

One way to decrease the length and size of trials is to power the study to detect changes in surrogate end points (e.g., a reduction in the absolute number of late-stage diagnoses versus controls) rather than mortality (*128*). Such trials are faster and require fewer participants to record enough events in a limited time frame. However, most health care systems, regulatory agencies, and guideline bodies still require evidence of reduced mortality before approving tests for marketing, reimbursement, or widespread use. Advice should be sought from the relevant agencies regarding acceptability of surrogate end points. Studies should be designed and powered, and end points chosen, on the basis of the objectives of the study (e.g., initial signal-finding trial versus technology validation versus confirmatory trial) and the intended circumstances of use. Early detection technology must generate the evidence that is required by regulators, advisory bodies, and payers for research to achieve clinical impact. Proper validation (*129*) and consideration of the pathway to implementation are crucial.

Interpreting trial results

Clinical trial results of early detection technologies should be interpreted taking into consideration spectrum and lead-time biases. Spectrum bias arises when tests are assessed in a population that does not reflect the intended target population (*130*). For example, comparing a study population with established and advanced disease to a healthy control population (often young and without other chronic diseases that increase variability in the general population) can confound the specificity of the test. Spectrum bias also arises when a test is developed using an at-risk population (e.g., heavy smokers) with high disease incidence but is intended for use in the general population (with lower incidence). Such a test will lose sensitivity and even specificity in the real-world target population, which has lower prevalence of disease and other confounders. This can cause false positives and even overdiagnosis.

Lead-time bias describes the time from early detection of disease to clinical presentation of signs and symptoms (when diagnosis would otherwise have taken place) (*131*, *132*). This makes survival seem longer when you detect cancer earlier by artificially moving the starting block back in time, even if early detection did not affect the point at which the individual died.

Spectrum bias can be addressed by validating markers and tests in populations that appropriately represent the population of intended use of the test. Lead-time bias is a more complex issue; currently, the method to address this bias is to conduct a trial designed to assess impact on mortality (e.g., whether there are fewer deaths overall in the screened group than in the unscreened group), however this then leads to the challenge of huge sample size and cost, as discussed above. This challenge can be addressed through careful study design. Dedicated experts in screening and diagnostic methodology must be involved, and the intended target audience for the results (e.g., regulatory and guideline-developing bodies) must be consulted when designing trials to evaluate early detection approaches. If we are to increase the glacial pace at which new early detection and screening approaches are evaluated and reach the clinic, a rethink of the evidence threshold for adoption is required. For example, success could be assessed on the basis of an absolute reduction in late-stage diagnoses (or other well-validated surrogate outcomes), with mortality data then gathered after implementation.

Conclusions

Early detection of cancer has the potential to transform patient survival and is increasingly recognized as an area of unmet need by the public, patients, policy-makers, and research funders. A sustained effort will be required to find practical, long-term solutions for many of the challenges we have described in this Review. We have suggested a framework that we believe will meaningfully accelerate progress (Fig. 6). Several contextual issues must also be carefully considered to maximize the translation of early detection research into clinical impact.

Funding programs targeting early detection have been set up by some funders of academic cancer research, such as the US National Cancer Institute (*133*, *134*) and Cancer Research UK (*135*, *136*). However, the proportion of overall cancer research funding dedicated to early detection remains disproportionately low considering the potential health benefits. More must be done, particularly in supporting validation of markers and tests (*129*). Dedicated funding would also help attract early career researchers to the field and enable them to become established. The relatively long timelines of early detection research and test development necessitate a rethink about traditional fellowship and grant models of supporting and evaluating early career researchers to incentivize them to establish a career in this field.

Furthermore, the pharmaceutical industry has invested proportionally little in early detection compared with the billions spent on drug development, often because of a historical

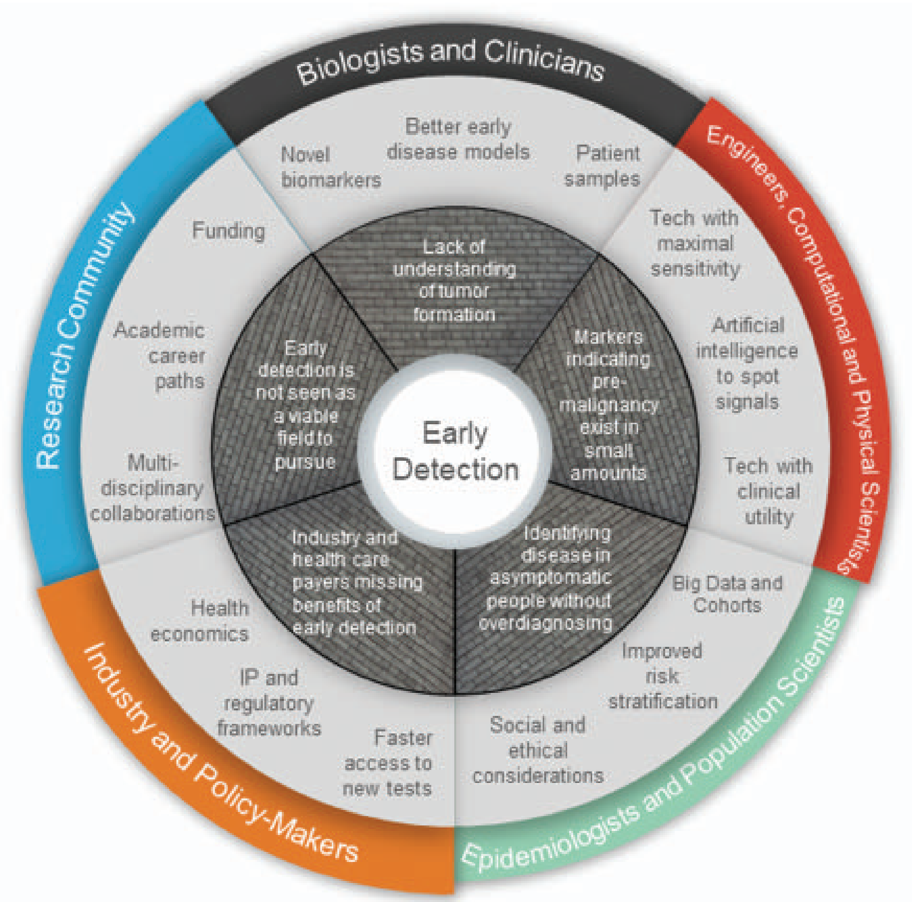


Fig. 6. Overcoming barriers to enable early detection. There are system-wide challenges (gray bricks) that must be tackled to reach the goal of earlier cancer detection. The multiple facets of these challenges require a diverse set of approaches and enablers (light gray) and communities (colored outer segments) to overcome them.

perception of an unattractive business model. However, there now appears to be an inflection (*137*) whereby investors and large corporations are increasingly willing to invest in this space (*138*, *139*). This willingness may be due to a growing realization that early detection will change the business model for cancer treatment.

An interdisciplinary culture is essential to early detection research and development, which inherently needs a convergence of biological understanding, clinical insight, technology innovation, data science, risk stratification, and health systems research. In the absence of any one of these essential components, the goal of transforming cancer survival cannot be realized. The implementation of interdisciplinarity can be fostered by research funders.

To have a meaningful impact on survival, early detection must be integrated into health care systems and must lead to evidence-based early interventions, either to prevent progression or to cure cancer. Lastly, and crucially, researchers must keep in mind that early detection should be accessible to all according to need, must not exacerbate health inequities, and must seek to do no harm (minimizing overdiagnosis and overtreatment). With the ever-increasing depth of biological insight and pace of technological innovation, we are at the tipping point for early cancer detection research and its translation to the ultimate objective of early curative interventions and increased cancer survival.

REFERENCES AND NOTES

1. H. Sung *et al*., Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021). doi: 10.3322/caac.21660; pmid: 33538338
2. World Health Organization, Fact sheet: Cancer; www.who.int/news-room/fact-sheets/detail/cancer.
3. Office for National Statistics UK, Cancer survival by stage at diagnosis for England (experimental statistics): Adults diagnosed 2012, 2013 and 2014 and followed up to 2015, 10 June 2016; www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancersurvivalbystageatdiagnosisforenglandexperimentalstatistics/adultsdiagnosed20122013and2014andfollowedupto2015.
4. R. L. Siegel *et al*., Colorectal cancer statistics, 2017. *CA Cancer J. Clin.* **67**, 177–193 (2017). doi: 10.3322/caac.21395; pmid: 28248415
5. E. Jedy-Agba, V. McCormack, C. Adebamowo, I. Dos-Santos-Silva, Stage at diagnosis of breast cancer in sub-Saharan Africa: A systematic review and meta-analysis. *Lancet Glob. Health* **4**, e923–e935 (2016). doi: 10.1016/S2214-109X(16)30259-5; pmid: 27855871
6. E. Jedy-Agba *et al*., Determinants of stage at diagnosis of breast cancer in Nigerian women: Sociodemographic, breast cancer awareness, health care access and clinical factors. *Cancer Causes Control* **28**, 685–697 (2017). doi: 10.1007/s10552-017-0894-y; pmid: 28447308
7. A. M. Nelson, D. A. Milner, T. R. Rebbeck, Y. Iliyasu, Oncologic care and pathology resources in Africa: Survey and recommendations. *J. Clin. Oncol.* **34**, 20–26 (2016). doi: 10.1200/JCO.2015.61.9767; pmid: 26578619
8. I. O. Morhason-Bello *et al*., Challenges and opportunities in cancer control in Africa: A perspective from the African Organisation for Research and Training in Cancer. *Lancet Oncol.* **14**, e142–e151 (2013). doi: 10.1016/S1470-2045(12)70482-5; pmid: 23561745
9. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011). doi: 10.1016/j.cell.2011.02.013; pmid: 21376230
10. N. Pashayan, P. D. P. Pharoah, The challenge of early detection in cancer. *Science* **368**, 589–590 (2020). doi: 10.1126/science.aaz2078; pmid: 32381710
11. R. A. Kyle *et al*., Prevalence of monoclonal gammopathy of undetermined significance. *N. Engl. J. Med.* **354**, 1362–1369 (2006). doi: 10.1056/NEJMoa054494; pmid: 16571879
12. Y. Peters *et al*., Barrett oesophagus. *Nat. Rev. Dis. Primers* **5**, 35 (2019). doi: 10.1038/s41572-019-0086-z; pmid: 31123267
13. K. M. Haigis, K. Cichowski, S. J. Elledge, Tissue-specificity in cancer: The rule, not the exception. *Science* **363**, 1150–1151 (2019). doi: 10.1126/science.aaw3472; pmid: 30872507
14. M. Fane, A. T. Weeraratna, How the ageing microenvironment influences tumour progression. *Nat. Rev. Cancer* **20**, 89–106 (2020). doi: 10.1038/s41568-019-0222-9; pmid: 31836838
15. C. J. Watson *et al*., The evolutionary dynamics and fitness landscape of clonal hematopoiesis. *Science* **367**, 1449–1454 (2020). doi: 10.1126/science.aay9333; pmid: 32217721
16. S. Manier *et al*., Genomic complexity of multiple myeloma and its clinical implications. *Nat. Rev. Clin. Oncol.* **14**, 100–113 (2017). doi: 10.1038/nrclinonc.2016.122; pmid: 27531699
17. M.-V. Mateos *et al*., Lenalidomide plus dexamethasone for high-risk smoldering multiple myeloma. *N. Engl. J. Med.* **369**, 438–447 (2013). doi: 10.1056/NEJMoa1300439; pmid: 23902483
18. H. Gonzalez, C. Hagerling, Z. Werb, Roles of the immune system in cancer: From tumor initiation to metastatic progression. *Genes Dev.* **32**, 1267–1284 (2018). doi: 10.1101/gad.314617.118; pmid: 30275043
19. T. J. Honkanen *et al*., Prognostic and predictive role of spatially positioned tumour infiltrating lymphocytes in metastatic HER2 positive breast cancer treated with trastuzumab. *Sci. Rep.* **7**, 18027 (2017). doi: 10.1038/s41598-017-18266-1; pmid: 29269742
20. I. Kirsch, M. Vignali, H. Robins, T-cell receptor profiling in cancer. *Mol. Oncol.* **9**, 2063–2070 (2015). doi: 10.1016/j.molonc.2015.09.003; pmid: 26404496
21. O. J. Finn, Immune response as a biomarker for cancer detection and a lot more. *N. Engl. J. Med.* **353**, 1288–1290 (2005). doi: 10.1056/NEJMe058157; pmid: 16177255
22. I. K. Macdonald, C. B. Parsy-Kowalska, C. J. Chapman, Autoantibodies: Opportunities for early cancer detection. *Trends Cancer* **3**, 198–213 (2017). doi: 10.1016/j.trecan.2017.02.003; pmid: 28718432
23. D. Beshnova *et al*., De novo prediction of cancer-associated T cell receptors for noninvasive cancer detection. *Sci. Transl. Med.* **12**, eaaz3738 (2020). doi: 10.1126/scitranslmed.aaz3738; pmid: 32817363
24. M. DuPage, T. Jacks, Genetically engineered mouse models of cancer reveal new insights about the antitumor immune response. *Curr. Opin. Immunol.* **25**, 192–199 (2013). doi: 10.1016/j.coi.2013.02.005; pmid: 23465466
25. B. Olson, Y. Li, Y. Lin, E. T. Liu, A. Patnaik, Mouse models for cancer immunotherapy research. *Cancer Discov.* **8**, 1358–1365 (2018). doi: 10.1158/2159-8290.CD-18-0044; pmid: 30309862
26. D. Bedognetti *et al*., Toward a comprehensive view of cancer immune responsiveness: A synopsis from the SITC workshop. *J. Immunother. Cancer* **7**, 131 (2019). doi: 10.1186/s40425-019-0602-4; pmid: 31113486
27. A. K. Palucka, L. M. Coussens, The basis of oncoimmunology. *Cell* **164**, 1233–1247 (2016). doi: 10.1016/j.cell.2016.01.049; pmid: 26967289
28. D. Hanahan, E. F. Wagner, R. D. Palmiter, The origins of oncomice: A history of the first transgenic mice genetically engineered to develop cancer. *Genes Dev.* **21**, 2258–2270 (2007). doi: 10.1101/gad.1583307; pmid: 17875663
29. D. Hanahan, R. A. Weinberg, The hallmarks of cancer. *Cell* **100**, 57–70 (2000). doi: 10.1016/S0092-8674(00)81683-9; pmid: 10647931
30. F. McCormick, Signalling networks that cause cancer. *Trends Cell Biol.* **9**, M53–M56 (1999). doi: 10.1016/S0962-8924(99)01668-2; pmid: 10611683
31. D. Hanahan, L. M. Coussens, Accessories to the crime: Functions of cells recruited to the tumor microenvironment. *Cancer Cell* **21**, 309–322 (2012). doi: 10.1016/j.ccr.2012.02.022; pmid: 22439926
32. N. Oshimori, D. Oristian, E. Fuchs, TGF-β promotes heterogeneity and drug resistance in squamous cell carcinoma. *Cell* **160**, 963–976 (2015). doi: 10.1016/j.cell.2015.01.043; pmid: 25723170
33. S. R. Hingorani *et al*., Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell* **4**, 437–450 (2003). doi: 10.1016/S1535-6108(03)00309-X; pmid: 14706336
34. K. M. Haigis *et al*., Differential effects of oncogenic K-Ras and N-Ras on proliferation, differentiation and tumor progression in the colon. *Nat. Genet.* **40**, 600–608 (2008). doi: 10.1038/ng.115; pmid: 18372904
35. E. L. Jackson *et al*., Analysis of lung tumor initiation and progression using conditional expression of oncogenic *K-ras*. *Genes Dev.* **15**, 3243–3248 (2001). doi: 10.1101/gad.943001; pmid: 11751630
36. A. L. Dooley *et al*., Nuclear factor I/B is an oncogene in small cell lung cancer. *Genes Dev.* **25**, 1470–1475 (2011). doi: 10.1101/gad.2046711; pmid: 21764851
37. S. C. Williamson *et al*., Vasculogenic mimicry in small cell lung cancer. *Nat. Commun.* **7**, 13322 (2016). doi: 10.1038/ncomms13322; pmid: 27827359
38. A. Lallo, M. W. Schenk, K. K. Frese, F. Blackhall, C. Dive, Circulating tumor cells and CDX models as a tool for preclinical drug development. *Transl. Lung Cancer Res.* **6**, 397–408 (2017). doi: 10.21037/tlcr.2017.08.01; pmid: 28904884
39. M. Breitenbach, J. Hoffmann, Editorial: Cancer models. *Front. Oncol.* **8**, 401 (2018). doi: 10.3389/fonc.2018.00401; pmid: 30338241
40. T. R. Rebbeck *et al*., Precision prevention and early detection of cancer: Fundamental principles. *Cancer Discov.* **8**, 803–811 (2018). doi: 10.1158/2159-8290.CD-17-1415; pmid: 29907587
41. M. Jansen, M. Banks, Early detection and risk stratification of gastric cancer are likely to be refined with biopsies targeted through high-resolution-enhanced imaging. *Gut* **69**, 1710–1711 (2020). doi: 10.1136/gutjnl-2019-319921; pmid: 31611299
42. A. Lee *et al*., BOADICEA: A comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genet. Med.* **21**, 1708–1718 (2019). doi: 10.1038/s41436-018-0406-9; pmid: 30643217
43. J. Tyrer, S. W. Duffy, J. Cuzick, A breast cancer prediction model incorporating familial and personal risk factors. *Stat. Med.* **23**, 1111–1130 (2004). doi: 10.1002/sim.1668; pmid: 15057881
44. S. A. Lambert, G. Abraham, M. Inouye, Towards clinical utility of polygenic risk scores. *Hum. Mol. Genet.* **28**, R133–R142 (2019). doi: 10.1093/hmg/ddz187; pmid: 31363735
45. E. W. Steyerberg *et al*., Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology* **21**, 128–138 (2010). doi: 10.1097/EDE.0b013e3181c30fb2; pmid: 20010215
46. A. Bleyer, H. G. Welch, Effect of three decades of screening mammography on breast-cancer incidence. *N. Engl. J. Med.* **367**, 1998–2005 (2012). doi: 10.1056/NEJMoa1206809; pmid: 23171096
47. A. C. Perkins, E. N. Skinner, A review of the current cervical cancer screening guidelines. *N. C. Med. J.* **77**, 420–422 (2016). doi: 10.18043/ncm.77.6.420; pmid: 27864494
48. R. Nishihara *et al*., Long-term colorectal-cancer incidence and mortality after lower endoscopy. *N. Engl. J. Med.* **369**, 1095–1105 (2013). doi: 10.1056/NEJMoa1301969; pmid: 24047059
49. K. L. Huang *et al*., Effects of low-dose computed tomography on lung cancer screening: A systematic review, meta-analysis, and trial sequential analysis. *BMC Pulm. Med.* **19**, 126 (2019). doi: 10.1186/s12890-019-0883-x; pmid: 31296196
50. M. Triplette, J. H. Thayer, S. N. Pipavath, K. Crothers, Poor uptake of lung cancer screening: Opportunities for improvement. *J. Am. Coll. Radiol.* **16**, 446–450 (2019). doi: 10.1016/j.jacr.2018.12.018; pmid: 30777648
51. J. Shreffler, M. R. Huecker, "Diagnostic Testing Accuracy: Sensitivity, Specificity, Predictive Values and Likelihood Ratios," StatPearls (StatPearls Publishing, 2021); www.ncbi.nlm.nih.gov/books/NBK557491/.
52. A. M. Molinaro, Diagnostic tests: How to estimate the positive predictive value. *Neurooncol. Pract.* **2**, 162–166 (2015). doi: 10.1093/nop/npv030; pmid: 31386059
53. D. Ilic, M. M. Neuberger, M. Djulbegovic, P. Dahm, Screening for prostate cancer. *Cochrane Database Syst. Rev.* **1**, CD004720 (2013). pmid: 23440794
54. D. C. Grossman *et al*., Screening for prostate cancer: US Preventive Services Task Force recommendation statement. *JAMA* **319**, 1901–1913 (2018). doi: 10.1001/jama.2018.3710; pmid: 29801017
55. U. Menon *et al*., Ovarian cancer population screening and mortality after long-term follow-up in the UK Collaborative

Trial of Ovarian Cancer Screening (UKCTOCS): A randomised controlled trial. *Lancet* **397**, 2182–2193 (2021). doi: 10.1016/S0140-6736(21)00731-5; pmid: 33991479

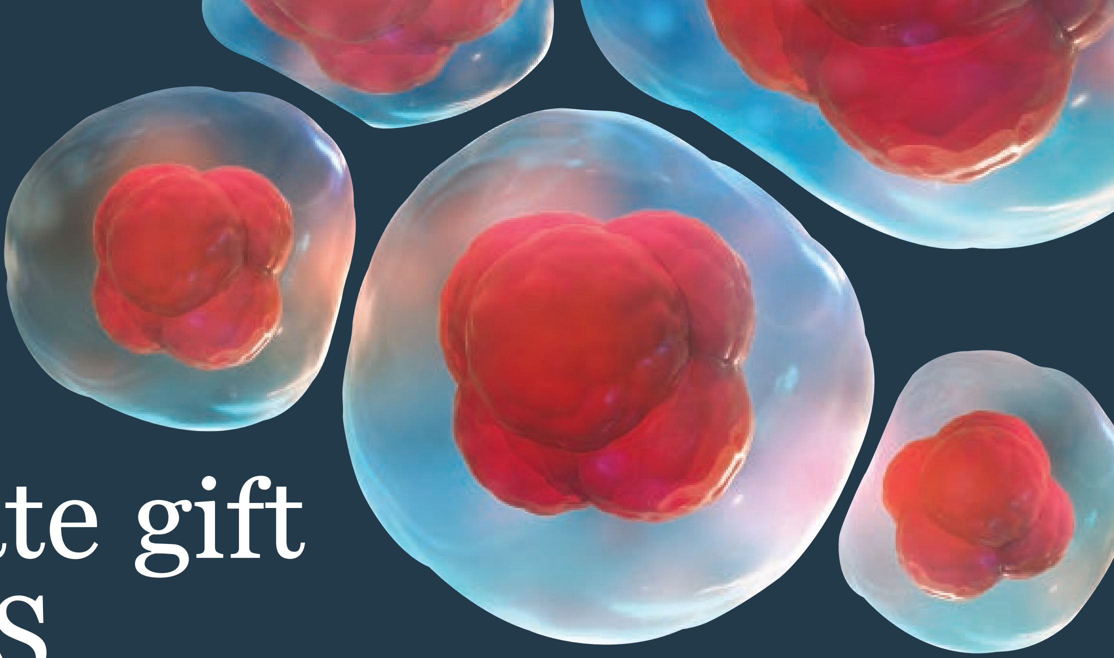
56. A. T. Boutin *et al*., Oncogenic *Kras* drives invasion and maintains metastases in colorectal cancer. *Genes Dev.* **31**, 370–382 (2017). doi: 10.1101/gad.293449.116; pmid: 28289141
57. M. P. di Magliano, C. D. Logsdon, Roles for KRAS in pancreatic tumor development and progression. *Gastroenterology* **144**, 1220–1229 (2013). doi: 10.1053/j.gastro.2013.01.071; pmid: 23622131
58. E. Heitzer, I. S. Haque, C. E. S. Roberts, M. R. Speicher, Current and future perspectives of liquid biopsies in genomics-driven oncology. *Nat. Rev. Genet.* **20**, 71–88 (2019). doi: 10.1038/s41576-018-0071-5; pmid: 30410101
59. S. X. Antoniou *et al*., The potential of breath analysis to improve outcome for patients with lung cancer. *J. Breath Res.* **13**, 034002 (2019). doi: 10.1088/1752-7163/ab0bee; pmid: 30822771
60. V. A. Adalsteinsson *et al*., Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat. Commun.* **8**, 1324 (2017). doi: 10.1038/s41467-017-00965-y; pmid: 29109393
61. International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004). doi: 10.1038/nature03001; pmid: 15496913
62. D. R. Bentley *et al*., Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008). doi: 10.1038/nature07517; pmid: 18987734
63. S. J. Jones *et al*., Evolution of an adenocarcinoma in response to selection by targeted kinase inhibitors. *Genome Biol.* **11**, R82 (2010). doi: 10.1186/gb-2010-11-8-r82; pmid: 20696054
64. T. J. Hudson *et al*., International network of cancer genome projects. *Nature* **464**, 993–998 (2010). doi: 10.1038/nature08987; pmid: 20393554
65. C. Curtis *et al*., The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012). doi: 10.1038/nature10983; pmid: 22522925
66. N. Wan *et al*., Machine learning enables detection of early-stage colorectal cancer by whole-genome sequencing of plasma cell-free DNA. *BMC Cancer* **19**, 832 (2019). doi: 10.1186/s12885-019-6003-8; pmid: 31443703
67. 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
68. A. P. Feinberg, The key role of epigenetics in human disease prevention and mitigation. *N. Engl. J. Med.* **378**, 1323–1334 (2018). doi: 10.1056/NEJMra1402513; pmid: 29617578
69. Y. Long, X. Wang, D. T. Youmans, T. R. Cech, How do lncRNAs regulate transcription? *Sci. Adv.* **3**, eaao2110 (2017). doi: 10.1126/sciadv.aao2110; pmid: 28959731
70. M. Tahiliani *et al*., Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009). doi: 10.1126/science.1170116; pmid: 19372391
71. M. C. Liu, G. R. Oxnard, E. A. Klein, C. Swanton, M. V. Seiden; CCGA Consortium, Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann. Oncol.* **31**, 745–759 (2020). doi: 10.1016/j.annonc.2020.02.011; pmid: 33506766
72. S. Cristiano *et al*., Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature* **570**, 385–389 (2019). doi: 10.1038/s41586-019-1272-6; pmid: 31142840
73. A. Carlsson *et al*., Circulating tumor microemboli diagnostics for patients with non-small-cell lung cancer. *J. Thorac. Oncol.* **9**, 1111–1119 (2014). doi: 10.1097/JTO.0000000000000235; pmid: 25157764
74. S. H. Jalalian, M. Ramezani, S. A. Jalalian, K. Abnous, S. M. Taghdisi, Exosomes, new biomarkers in early cancer detection. *Anal. Biochem.* **571**, 1–13 (2019). doi: 10.1016/j.ab.2019.02.013; pmid: 30776327
75. C. E. Gast *et al*., Cell fusion potentiates tumor heterogeneity and reveals circulating hybrid cells that correlate with stage and survival. *Sci. Adv.* **4**, eaat7828 (2018). doi: 10.1126/sciadv.aat7828; pmid: 30214939
76. A. Srivastava, D. J. Creek, Discovery and validation of clinical biomarkers of cancer: A review combining metabolomics and proteomics. *Proteomics* **19**, 1700448 (2019). doi: 10.1002/pmic.201700448; pmid: 30353665
77. H. T. Tan, Y. H. Lee, M. C. Chung, Cancer proteomics. *Mass Spectrom. Rev.* **31**, 583–605 (2012). doi: 10.1002/mas.20356; pmid: 22422534
78. M. Bhardwaj *et al*., Multiplex quantitation of 270 plasma protein markers to identify a signature for early detection of colorectal cancer. *Eur. J. Cancer* **127**, 30–40 (2020). doi: 10.1016/j.ejca.2019.11.021; pmid: 31972396
79. M. Bhardwaj, A. Gies, S. Werner, P. Schrotz-King, H. Brenner, Blood-based protein signatures for early detection of colorectal cancer: A systematic review. *Clin. Transl. Gastroenterol.* **8**, e128 (2017). doi: 10.1038/ctg.2017.53; pmid: 29189767
80. J. D. Cohen *et al*., Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **359**, 926–930 (2018). doi: 10.1126/science.aar3247; pmid: 29348365
81. V. Gopalakrishnan *et al*., Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science* **359**, 97–103 (2018). doi: 10.1126/science.aan4236; pmid: 29097493
82. V. Matson *et al*., The commensal microbiome is associated with anti–PD-1 efficacy in metastatic melanoma patients. *Science* **359**, 104–108 (2018). doi: 10.1126/science.aao3290; pmid: 29302014
83. B. Routy *et al*., Gut microbiome influences efficacy of PD-1–based immunotherapy against epithelial tumors. *Science* **359**, 91–97 (2018). doi: 10.1126/science.aan3706; pmid: 29097494
84. B. A. Helmink, M. A. W. Khan, A. Hermann, V. Gopalakrishnan, J. A. Wargo, The microbiome, cancer, and cancer therapy. *Nat. Med.* **25**, 377–388 (2019). doi: 10.1038/s41591-019-0377-7; pmid: 30842679
85. S. U. Springer *et al*., Non-invasive detection of urothelial cancer through the analysis of driver gene mutations and aneuploidy. *eLife* **7**, e32143 (2018). doi: 10.7554/eLife.32143; pmid: 29557778
86. Y. Wang *et al*., Evaluation of liquid from the Papanicolaou test and other liquid biopsies for the detection of endometrial and ovarian cancers. *Sci. Transl. Med.* **10**, eaap8793 (2018). doi: 10.1126/scitranslmed.aap8793; pmid: 29563323
87. Z. Hu *et al*., Quantitative evidence for early metastatic seeding in colorectal cancer. *Nat. Genet.* **51**, 1113–1122 (2019). doi: 10.1038/s41588-019-0423-x; pmid: 31209394
88. A. M. Lennon *et al*., Feasibility of blood testing combined with PET-CT to screen for cancer and guide intervention. *Science* **369**, eabb9601 (2020). doi: 10.1126/science.abb9601; pmid: 32345712
89. T. F. Imperiale *et al*., Multitarget stool DNA testing for colorectal-cancer screening. *N. Engl. J. Med.* **370**, 1287–1297 (2014). doi: 10.1056/NEJMoa1311194; pmid: 25006736
90. C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **1**, 206–215 (2019). doi: 10.1038/s42256-019-0048-x
91. J. Dockès, G. Varoquaux, J. B. Poline, Preventing dataset shift from breaking machine-learning biomarkers. *Gigascience* **10**, giab055 (2021). doi: 10.1093/gigascience/giab055; pmid: 34585237
92. C. L. Andaur Navarro *et al*., Risk of bias in studies on prediction models developed using supervised machine learning techniques: Systematic review. *BMJ* **375**, n2281 (2021). pmid: 34670780
93. M. Nagendran *et al*., Artificial intelligence versus clinicians: Systematic review of design, reporting standards, and claims of deep learning studies. *BMJ* **368**, m689 (2020). doi: 10.1136/bmj.m689; pmid: 32213531
94. E. Christodoulou *et al*., A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J. Clin. Epidemiol.* **110**, 12–22 (2019). doi: 10.1016/j.jclinepi.2019.02.004; pmid: 30763612
95. V. Baeriswyl, G. Christofori, The angiogenic switch in carcinogenesis. *Semin. Cancer Biol.* **19**, 329–337 (2009). doi: 10.1016/j.semcancer.2009.05.003; pmid: 19482086
96. X. He *et al*., First in-vivo human imaging at 10.5T: Imaging the body at 447 MHz. *Magn. Reson. Med.* **84**, 289–303 (2020). doi: 10.1002/mrm.28131; pmid: 31846121
97. A. P. Soleimany, S. N. Bhatia, Activity-based diagnostics: An emerging paradigm for disease detection and monitoring. *Trends Mol. Med.* **26**, 450–468 (2020). doi: 10.1016/j.molmed.2020.01.013; pmid: 32359477
98. S. Slomovic, K. Pardee, J. J. Collins, Synthetic biology devices for in vitro and in vivo diagnostics. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 14429–14435 (2015). doi: 10.1073/pnas.1508521112; pmid: 26598662
99. C. N. Loynachan *et al*., Renal clearable catalytic gold nanoclusters for in vivo disease monitoring. *Nat. Nanotechnol.* **14**, 883–890 (2019). doi: 10.1038/s41565-019-0527-6; pmid: 31477801
100. T. Danino *et al*., Programmable probiotics for detection of cancer in urine. *Sci. Transl. Med.* **7**, 289ra84 (2015). doi: 10.1126/scitranslmed.aaa3519; pmid: 26019220
101. A. Aalipour *et al*., Engineered immune cells as highly sensitive cancer diagnostics. *Nat. Biotechnol.* **37**, 531–539 (2019). doi: 10.1038/s41587-019-0064-8; pmid: 30886438
102. A. F. Sarioglu *et al*., A microfluidic device for label-free, physical capture of circulating tumor cell clusters. *Nat. Methods* **12**, 685–691 (2015). doi: 10.1038/nmeth.3404; pmid: 25984697
103. P. Zhang *et al*., Ultrasensitive detection of circulating exosomes with a 3D-nanopatterned microfluidic chip. *Nat. Biomed. Eng.* **3**, 438–451 (2019). doi: 10.1038/s41551-019-0356-9; pmid: 31123323
104. S. S. Gambhir, T. J. Ge, O. Vermesh, R. Spitler, Toward achieving precision health. *Sci. Transl. Med.* **10**, eaao3612 (2018). doi: 10.1126/scitranslmed.aao3612; pmid: 29491186
105. H. U. Kauczor *et al*., ESR/ERS statement paper on lung cancer screening. *Eur. Radiol.* **30**, 3277–3294 (2020). doi: 10.1007/s00330-020-06727-7; pmid: 32052170
106. J. C. Gore, H. C. Manning, C. C. Quarles, K. W. Waddell, T. E. Yankeelov, Magnetic resonance in the era of molecular imaging of cancer. *Magn. Reson. Imaging* **29**, 587–600 (2011). doi: 10.1016/j.mri.2011.02.003; pmid: 21524870
107. J. V. Frangioni, New technologies for human cancer imaging. *J. Clin. Oncol.* **26**, 4012–4021 (2008). doi: 10.1200/JCO.2007.14.3065; pmid: 18711192
108. S. Surti *et al*., Performance of Philips Gemini TF PET/CT scanner with special consideration for its time-of-flight imaging capabilities. *J. Nucl. Med.* **48**, 471–480 (2007). pmid: 17332626
109. K. Golman, R. I. Zandt, M. Lerche, R. Pehrson, J. H. Ardenkjaer-Larsen, Metabolic imaging by hyperpolarized 13C magnetic resonance imaging for in vivo tumor diagnosis. *Cancer Res.* **66**, 10855–10860 (2006). doi: 10.1158/0008-5472.CAN-06-2564; pmid: 17108122
110. J. Kurhanewicz *et al*., Hyperpolarized ^{13}C MRI: Path to clinical translation in oncology. *Neoplasia* **21**, 1–16 (2019). doi: 10.1016/j.neo.2018.09.006; pmid: 30472500
111. H. U. Ahmed *et al*., Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): A paired validating confirmatory study. *Lancet* **389**, 815–822 (2017). doi: 10.1016/S0140-6736(16)32401-1; pmid: 28110982
112. J. Katzen, K. Dodelzon, A review of computer aided detection in mammography. *Clin. Imaging* **52**, 305–309 (2018). doi: 10.1016/j.clinimag.2018.08.014; pmid: 30216858
113. H. J. Yoon *et al*., Deciphering the tumor microenvironment through radiomics in non-small cell lung cancer: Correlation with immune profiles. *PLOS ONE* **15**, e0231227 (2020). doi: 10.1371/journal.pone.0231227; pmid: 32251447
114. S. M. McKinney *et al*., International evaluation of an AI system for breast cancer screening. *Nature* **577**, 89–94 (2020). doi: 10.1038/s41586-019-1799-6; pmid: 31894144
115. M. Bahl *et al*., High-risk breast lesions: A machine learning model to predict pathologic upgrade and reduce unnecessary surgical excision. *Radiology* **286**, 810–818 (2018). doi: 10.1148/radiol.2017170549; pmid: 29039725
116. B. Haibe-Kains *et al*., Transparency and reproducibility in artificial intelligence. *Nature* **586**, E14–E16 (2020). doi: 10.1038/s41586-020-2766-y; pmid: 33057217
117. I. Steinberg *et al*., Photoacoustic clinical imaging. *Photoacoustics* **14**, 77–98 (2019). doi: 10.1016/j.pacs.2019.05.001; pmid: 31293884
118. W. B. Nagengast *et al*., Near-infrared fluorescence molecular endoscopy detects dysplastic oesophageal lesions using topical and systemic tracer of vascular endothelial growth factor A. *Gut* **68**, 7–10 (2019). doi: 10.1136/gutjnl-2017-314953; pmid: 29247063
119. J. Burggraaf *et al*., Detection of colorectal polyps in humans using an intravenously administered fluorescent peptide targeted against c-Met. *Nat. Med.* **21**, 955–961 (2015). doi: 10.1038/nm.3641; pmid: 26168295
120. K. Bera, K. A. Schalper, D. L. Rimm, V. Velcheti, A. Madabhushi, Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. *Nat. Rev. Clin. Oncol.* **16**, 703–715 (2019). doi: 10.1038/s41571-019-0252-y; pmid: 31399699

121. M. K. K. Niazi, A. V. Parwani, M. N. Gurcan, Digital pathology and artificial intelligence. *Lancet Oncol.* **20**, e253–e261 (2019). doi: 10.1016/S1470-2045(19)30154-8; pmid: 31044723
122. P. Hu, M. Zelen, Planning of randomized early detection trials. *Stat. Methods Med. Res.* **13**, 491–506 (2004). doi: 10.1191/0962280204sm379ra; pmid: 15587435
123. D. R. Aberle *et al*., Reduced lung-cancer mortality with low-dose computed tomographic screening. *N. Engl. J. Med.* **365**, 395–409 (2011). doi: 10.1056/NEJMoa1102873; pmid: 21714641
124. H. J. de Koning *et al*., Reduced lung-cancer mortality with volume CT screening in a randomized trial. *N. Engl. J. Med.* **382**, 503–513 (2020). doi: 10.1056/NEJMoa1911793; pmid: 31995683
125. U. Menon *et al*., Sensitivity and specificity of multimodal and ultrasound screening for ovarian cancer, and stage distribution of detected cancers: Results of the prevalence screen of the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *Lancet Oncol.* **10**, 327–340 (2009). doi: 10.1016/S1470-2045(09)70026-9; pmid: 19282241
126. J. Hugosson *et al*., A 16-yr Follow-up of the European Randomized study of Screening for Prostate Cancer. *Eur. Urol.* **76**, 43–51 (2019). doi: 10.1016/j.eururo.2019.02.009; pmid: 30824296
127. L. Highfield *et al*., A non-randomized controlled stepped wedge trial to evaluate the effectiveness of a multi-level mammography intervention in improving appointment adherence in underserved women. *Implement. Sci.* **10**, 143 (2015). doi: 10.1186/s13012-015-0334-x; pmid: 26464110
128. J. Cuzick, F. H. Cafferty, R. Edwards, H. Møller, S. W. Duffy, Surrogate endpoints for cancer screening trials: General principles and an illustration using the UK Flexible Sigmoidoscopy Screening Trial. *J. Med. Screen.* **14**, 178–185 (2007). doi: 10.1258/096914107782912059; pmid: 18078562
129. G. Poste, Bring on the biomarkers. *Nature* **469**, 156–157 (2011). doi: 10.1038/469156a; pmid: 21228852
130. S. A. Mulherin, W. C. Miller, Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. *Ann. Intern. Med.* **137**, 598–602 (2002). doi: 10.7326/0003-4819-137-7-200210010-00011; pmid: 12353947
131. B. S. Kramer, The science of early detection. *Urol. Oncol.* **22**, 344–347 (2004). doi: 10.1016/j.urolonc.2003.04.001; pmid: 15283894
132. T. M. Andersson, M. J. Rutherford, K. Humphreys, Assessment of lead-time bias in estimates of relative survival for breast cancer. *Cancer Epidemiol.* **46**, 50–56 (2017). doi: 10.1016/j.canep.2016.12.004; pmid: 28027488
133. National Institutes of Health, National Cancer Institute, Early Detection Research Network; https://edrn.nci.nih.gov/.
134. National Institutes of Health, National Cancer Institute, Consortium for Molecular and Cellular Characterization of Screen-Detected Lesions; https://mcl.nci.nih.gov/.
135. Cancer Research UK, Early Detection and Diagnosis Research Committee; www.cancerresearchuk.org/funding-for-researchers/applying-for-funding/funding-committees/early-detection-and-diagnosis-committee.
136. Cancer Research UK, Early Detection and Diagnosis of Cancer Roadmap; www.cancerresearchuk.org/funding-for-researchers/research-opportunities-in-early-detection-and-diagnosis/early-detection-and-diagnosis-roadmap.
137. W. N. Hait, P. F. Lebowitz, Disease interception: Myths, mountains, and mole hills. *Cancer Prev. Res.* **9**, 635–637 (2016). doi: 10.1158/1940-6207.CAPR-16-0049; pmid: 27138792
138. D. Schenkel, S. Nambi, R. Blicker, C. Lin, Liquid Biopsy: Early Detection of a Huge Investment Opportunity; www.cowen.com/insights/liquid-biopsy-early-detection-of-a-huge-investment-opportunity/.
139. Grand View Research, "Cancer diagnostics market worth $249.6 billion by 2026," February 2019; www.grandviewresearch.com/press-release/global-cancer-diagnostics-market.
140. Office for National Statistics UK, Cancer survival in England: adult, stage at diagnosis and childhood - patients followed up to 2018, 12 August 2019; www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancersurvivalinengland/stageatdiagnosisandchildhoodpatientsfollowedupto2018.

ACKNOWLEDGMENTS

We dedicate this Review to our friend, colleague, and mentor Sanjiv "Sam" Gambhir. He showed us the path to early detection of cancer and taught us to always look to the horizon. We continue to walk that path in his memory. Paid editorial assistance was provided by F. Dunlevy, Axcience. We also thank A. Webb, Cancer Research UK, for editorial and logistical assistance during the preparation of this manuscript and for contributions to the figures. We acknowledge the Cancer Research UK Policy and Information Directorate for their assistance on UK cancer diagnosis and screening statistics. We also acknowledge the Early Detection, Prevention, and Infections Branch of the International Agency for Research on Cancer for their assistance on Asian cancer survival statistics. **Funding:** M.E. receives research support from the UK's National Institute of Health Research (NIHR) UCLH/UCL Biomedical Research Centre and was awarded NIHR Senior Investigator status in 2013. **Competing interests:** S.Ba. is a founder, board member, and shareholder of Cambridge Epigenetix Ltd. and a science partner and adviser to Ahren Innovation Capital. S.Bh. is a board member of Brown University and Vertex Pharmaceuticals; holds advisory roles at Chan Zuckerberg Biohub, Global Oncology Inc., Cristal Therapeutics, Glympse Bio, CEND Therapeutics, and Satellite Bio; consults for Moderna Therapeutics; and performs sponsored research for Johnson & Johnson. S.Bh. holds equity in Vertex Pharmaceuticals, Glympse Bio, CEND Therapeutics, and Satellite Bio. K.M.B. holds patents with GE Healthcare on some aspects of hyperpolarization technology and is a paid consultant for NVision Imaging Technologies GmbH and New Path Molecular Research Ltd. L.M.C. is a paid consultant for the Susan G. Komen Foundation, Cell Signaling Technologies, AbbVie Inc., and Shasqi Inc.; receives reagent and/or research support from Pharmacyclics Inc., Acerta Pharma LLC, Deciphera Pharmaceuticals LLC, Genentech Inc., Roche Glycart AG, Syndax Pharmaceuticals Inc., and NanoString Technologies; and is a member of the scientific advisory boards of Syndax Pharmaceuticals, Carisma Therapeutics, Zymeworks Inc., Verseau Therapeutics, Cytomix Therapeutics Inc., Alkermes Inc., Cell Signaling Technology, Genenta Science, and Kineta Inc.; and is a site lead for the AstraZeneca Partner of Choice Network. C.D. has received research funding and/or educational research grants from: AstraZeneca, Astex Pharmaceuticals, Bioven, Amgen, Carrick Therapeutics, Merck AG, Taiho Oncology, GSK, Bayer, Boehringer Ingelheim, Roche, BMS, Novartis, Celgene, Epigene Therapeutics Inc., Angle PLC, Menarini, Clearbridge Biomedics, Thermo Fisher Scientific, and Neomed Therapeutics. C.D. has received honoraria for consultancy and/or advisory boards from: Biocartis, Merck, AstraZeneca, GRAIL, and Boehringer Ingelheim. M.E. is an investigator, adviser, and/or lecturer to the following companies: STEBA Biotech, Sonacare Inc., and Angiodynamics Inc. R.C.F. is named on patents related to Cytosponge and associated assays, which has been licensed by the Medical Research Council to Covidien GI Solutions (now Medtronic). R.C.F. is a shareholder and consultant for Cyted Ltd., an early detection company. S.S.G. was a founder, board member, and stockholder of Earli Inc. P.K. is a named inventor on patents relating to liquid biopsy technology, which have been licensed to Epic Sciences by the Scripps Research Institute for commercial development. P.K. is a named inventor on patents relating to human performance assessment, which have been licensed to Cansera for commercial development. P.K. is a shareholder of and consultant to Epic Sciences and Cansera. P.K. receives research funding from Kite, Gilead, Amgen, and Novartis. P.K. is a member of the CRUK Early Detection committee.

10.1126/science.aay9040



1848
SOCIETY
AAAS

An estate gift to AAAS

Going all the way back to 1848, our founding year, the American Association for the Advancement of Science (AAAS) has been deeply committed to advancing science, engineering and innovation around the world for the benefit of all people.

Today, we are dedicated to advocating for science and scientific evidence to be fully and positively integrated into public policy and for the community to speak with one voice to advance science and engineering in the United States and around the world.

By making AAAS a beneficiary of your will, trust, retirement plan or life insurance policy, you will become a member of our 1848 Society and will help fuel our work on behalf of science and society – including publishing the world's most promising, innovative research in the *Science* family of journals and engaging in the issues that matter locally, nationally and around the world.

> *"As a teacher and instructor, I bear responsibility for the younger generations. If you have extra resources, concentrate them on organizations, like AAAS, that are doing work for all."*
>
> **—Prof. Elisabeth Ervin-Blankenheim, 1848 Society member**

If you intend to include AAAS in your estate plans, provide this information to your lawyer or financial adviser:

Legal Name: American Association for the Advancement of Science
Federal Tax ID Number: 53-0196568
Address: 1200 New York Avenue, NW, Washington, DC 20005

If you would like more information on making an estate gift to AAAS, cut out and return the form below or send an email to philanthropy@aaas.org. Additional details are also available online at **www.aaas.org/1848Society**.

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

✂ *cut here*

Yes, I would like more information about joining the AAAS 1848 Society.

PLEASE CONTACT ME AT:

1848
SOCIETY
AAAS

Name: ____________________

Address: ____________________

City: ____________ State: ______ Zip code: __________ Country: ________

Email: ____________________ Phone: ____________________

RETURN THIS FORM TO:
AAAS Office of Philanthropy and Strategic Partnerships • 1200 New York Avenue, NW • Washington, DC 20005 USA

RESEARCH ARTICLE SUMMARY

NEUROSCIENCE

Compartment-specific tuning of dendritic feature selectivity by intracellular Ca^{2+} release

Justin K. O'Hare, Kevin C. Gonzalez†, Stephanie A. Herrlinger†, Yusuke Hirabayashi†, Victoria L. Hewitt, Heike Blockus, Miklos Szoboszlay, Sebi V. Rolotti, Tristan C. Geiller, Adrian Negrean, Vikas Chelur, Franck Polleux*‡, Attila Losonczy*‡

INTRODUCTION: Synaptic plasticity, the process by which neurons adjust the strengths of their thousands of inputs, allows animals to adapt to their environment. Decades of research have established Ca^{2+} as a central mediator of synaptic plasticity. Historically, most investigations into the role of Ca^{2+} in plasticity have focused on its influx through voltage-gated channels that open when synaptic input depolarizes a neuron. However, a large body of in vitro research suggests that an alternate source of Ca^{2+} may also play a potent role in shaping plasticity: the endoplasmic reticulum (ER). The ER stores Ca^{2+} in vast quantities within a cell and can release this store in response to strong synaptic input through intracellular Ca^{2+} release (ICR). The ER is therefore poised to shape the magnitude and spatial distribution of Ca^{2+} during plasticity induction. Despite its potential role in synaptic plasticity, ICR has never been investigated in mammalian neurons in vivo.

RATIONALE: To test whether ICR participates in experience-dependent plasticity, we focused on pyramidal neurons of hippocampal area CA1 (CA1PNs). CA1PNs receive excitatory inputs from multiple afferent circuits, carrying complementary streams of information about an animal's environment that impinge onto distinct compartments of the CA1PN dendritic arbor. As an animal explores a new environment, CA1PNs integrate these inputs to form spatially tuned receptive fields known as place fields that manifest as a neuron firing when an animal occupies a particular location. Recent work has characterized an in vivo plasticity mechanism driving place field formation: behavioral time-scale plasticity (BTSP). BTSP is initiated by a large, prolonged dendritic depolarization (plateau potential) that ultimately potentiates synaptic inputs received during a seconds-long time window corresponding to the time of plateau onset.

Here, we used CA1PNs and BTSP as a model system to test whether ICR participates in the experience-dependent emergence of feature selectivity. We implemented a series of tools based on single-cell electroporation allowing us to (i) manipulate the cytosolic impact of ICR at single-cell resolution using conditional gene deletion, (ii) optogenetically induce place cells, and (iii) image somatic and dendritic Ca^{2+} dynamics simultaneously during spatial navigation.

RESULTS: The gene *Pdzd8* encodes a recently identified ER-mitochondrial tethering protein that, when deleted, leads to unrestricted ICR. We found that *Pdzd8* deletion in single adult CA1PNs in vivo substantially increased the level of spatial co-tuning observed in their apical dendrites relative to the soma of CA1 place cells, a phenomenon not observed in basal dendrites, which were already highly co-tuned with the soma in control CA1PNs. Maximizing ICR leads to more-stable retention of place cell spatial tuning over time and alters the integrative properties of their apical dendrites to shape output-level receptive fields.

CONCLUSION: ICR plays a key and previously uncharacterized role in shaping the dendritic integration properties of CA1PNs during the emergence of feature selectivity. Therefore, ICR cooperates with circuit-level architecture in vivo to promote the emergence of behaviorally relevant forms of plasticity in a compartment-specific manner. ■

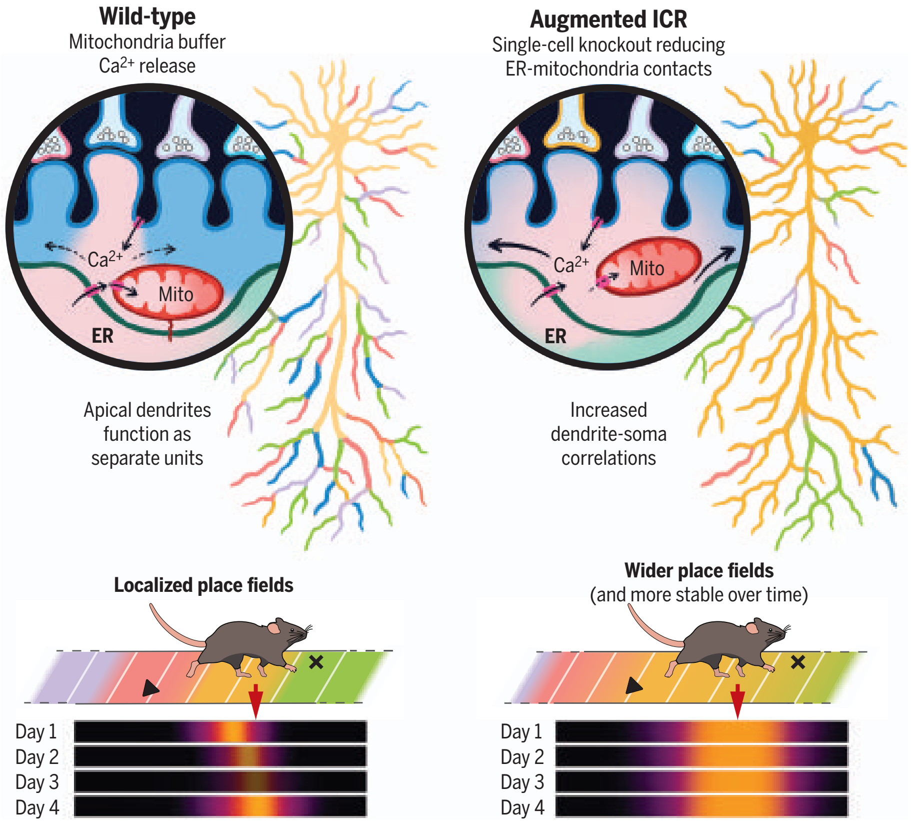


ILLUSTRATION BY MATTEO FARINELLA

ICR shapes receptive fields important for spatial navigation. We tested the role of ICR in neuronal feature selectivity by genetically ablating contact sites between ER and mitochondria (Mito) in single pyramidal neurons of hippocampal area CA1 in vivo. Increasing the impact of ICR on cytoplasmic Ca^{2+} altered integrative properties in apical dendrites, widening and stabilizing "place fields" over time.

The list of author affiliations is available in the full article online.
***Corresponding author. Email: al2856@columbia.edu (A.L.); fp2304@columbia.edu (F.P.)**
†These authors contributed equally to this work.
‡These authors contributed equally to this work and are co-senior authors.

Cite this article as J. K. O'Hare *et al.*, *Science* 375, eabm1670 (2022). DOI: 10.1126/science.abm1670

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abm1670

RESEARCH ARTICLE

NEUROSCIENCE

Compartment-specific tuning of dendritic feature selectivity by intracellular Ca^{2+} release

Justin K. O'Hare[1,2], Kevin C. Gonzalez[1,2]†, Stephanie A. Herrlinger[1,2]†, Yusuke Hirabayashi[3]†, Victoria L. Hewitt[1,2], Heike Blockus[1,2], Miklos Szoboszlay[1,2], Sebi V. Rolotti[1,2], Tristan C. Geiller[1,2], Adrian Negrean[1,2], Vikas Chelur[1], Franck Polleux[1,2,4]*‡, Attila Losonczy[1,2,4]*‡

Dendritic calcium signaling is central to neural plasticity mechanisms that allow animals to adapt to the environment. Intracellular calcium release (ICR) from the endoplasmic reticulum has long been thought to shape these mechanisms. However, ICR has not been investigated in mammalian neurons in vivo. We combined electroporation of single CA1 pyramidal neurons, simultaneous imaging of dendritic and somatic activity during spatial navigation, optogenetic place field induction, and acute genetic augmentation of ICR cytosolic impact to reveal that ICR supports the establishment of dendritic feature selectivity and shapes integrative properties determining output-level receptive fields. This role for ICR was more prominent in apical than in basal dendrites. Thus, ICR cooperates with circuit-level architecture in vivo to promote the emergence of behaviorally relevant plasticity in a compartment-specific manner.

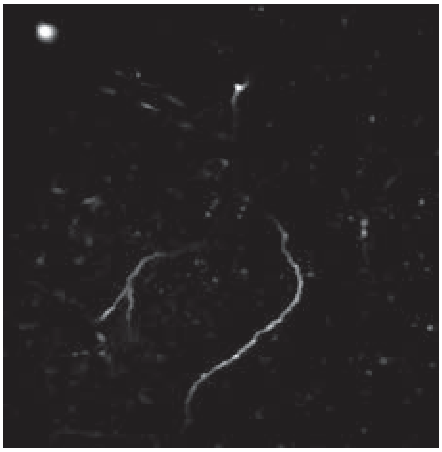
Learning occurs when experience-driven neuronal activity patterns induce changes in synaptic strengths, thereby altering how future information propagates through neuronal circuits. This process, known as synaptic plasticity, is a fundamental way in which animals adapt to the environment and yet it remains enigmatic. At the cellular level, Ca^{2+} helps to transduce specific patterns of synaptic activation into long-lasting changes in synaptic efficacy (*1*). The magnitude and spatiotemporal patterning of cytosolic Ca^{2+} are therefore critical in determining which synapses will undergo plasticity and the extent to which they will do so. Most studies of the mechanisms regulating dendritic Ca^{2+} dynamics have focused on the role of voltage-gated channels mediating Ca^{2+} conductances at the plasma membrane (*2*–*4*). However, another regulator of intracellular Ca^{2+} dynamics is the main internal Ca^{2+} store, the endoplasmic reticulum (ER) (*5*), which pervades the dendritic arbor (*6*–*8*), where it sequesters nearly all Ca^{2+} within a neuron. The ER can release its highly concentrated store of Ca^{2+} in an activity-dependent manner, significantly amplifying the impact of Ca^{2+} influx from the extracellular space (*5*, *9*–*14*).

Although much attention has been paid to the potential role for intracellular Ca^{2+} release (ICR) in plasticity in numerous in vitro studies (*11*–*21*), and recently in *Drosophila* (*22*), this powerful intracellular amplificatory process remains unexplored in mammalian neurons in vivo because of the lack of tools to precisely and effectively manipulate ICR in awake, behaving vertebrates. Such an approach would require not only cellular and pre- versus postsynaptic specificity but also the ability to influence ICR across each of its two canonical release mechanisms (*5*, *23*). To overcome this limitation, we focused on *Pdzd8*, a gene that encodes a tethering protein that brings ER and mitochondria into direct contact, thereby enabling mitochondria to buffer a substantial fraction of ER-released Ca^{2+} (*9*). Reduction of *Pdzd8* expression allows ~50% more Ca^{2+} to escape into the cytosol upon synaptically driven ICR in dendrites while leaving functional and morphological properties of the ER and mitochondria intact (*9*). Therefore, *Pdzd8*-mediated ER-mitochondria contacts present an avenue to manipulate ICR in vivo across its multiple release pathways.

Here, we leveraged *Pdzd8* deletion as a gain-of-function tool to augment the cytosolic impact of ICR. We deployed this molecular-genetic tool in pyramidal neurons of hippocampal area CA1 (CA1PNs). The CA1PN is a rich model system with which to investigate the general principles of experience-dependent synaptic plasticity. It receives diverse, circuit-specific excitatory inputs transmitting complementary streams of information relevant to an animal's environment, and these inputs target dendritic compartments that are morphologically and biophysically distinct from one another (*24*–*29*). As an animal explores its environment, CA1PNs integrate these diverse inputs to generate spatially tuned receptive fields, known as "place fields" (PFs), which manifest as the cell selectively firing when the animal occupies a particular location within the environment and collectively form a comprehensive map of an animal's environment (*30*, *31*). Recent work has established a plasticity mechanism, known as behavioral time-scale plasticity (BTSP) (*32*–*34*), which is sufficient to drive rapid PF formation in CA1PNs. BTSP is driven by large, prolonged dendritic depolarization events, known as plateau potentials, which potentiate dendritic spines receiving presynaptic input within a seconds-long temporal window surrounding plateau potential onset. We leveraged the well-characterized functional-anatomical complexity of CA1PNs and the robust in vivo plasticity mechanism BTSP to test the role of ICR in synaptic plasticity mechanisms supporting dendritic and cellular feature selectivity.

Single-cell genetic deletion and in vivo subcellular imaging in hippocampal area CA1

We simultaneously monitored the activity of dendrites and their cognate somata in CA1PNs with either normal or augmented ICR during spatial navigation. Using in vivo single-cell electroporation (see the materials and methods), we introduced pCAG-Cre, pCAG-FLEX-jGCaMP7b, and pCaMKII-bReaChes-mRuby3 to single CA1PNs in anesthetized adult mice (Fig. 1A, left and middle). We targeted single cells in a *Pdzd8* conditional knockout mouse line $Pdzd8^{F/F}$ (fig. S1, A and B) and in wild-type (WT) control mice to generate single *Pdzd8*-knockout (KO) (augmented ICR) and WT cells (Fig. 1A, right) expressing a high-baseline



Movie 1. Somatic (left) and apical dendritic (right) imaging planes of example CA1PN shown in Fig. 1B (right) displayed side by side during a PF induction lap. jGCaMP7b response to a single, 1-s LED pulse is shown. Frames are motion corrected and shown at actual speed. Data were acquired using two-photon microscopy with a 40× objective lens at 1.0× optical zoom.

---

[1] Department of Neuroscience, Columbia University, New York, NY 10027, USA. [2] Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY 10027, USA. [3] Department of Chemistry and Biotechnology, School of Engineering, The University of Tokyo, Tokyo 113-8656, Japan. [4] Kavli Institute for Brain Science, Columbia University, New York, NY 10027, USA.
*Corresponding author. Email: al2856@columbia.edu (A.L.); fp2304@columbia.edu (F.P.)
†These authors contributed equally to this work. ‡These authors contributed equally to this work and are co-senior authors.

Ca^{2+} indicator (jGCaMP7b) and a fluorescently tagged excitatory opsin (bReaChes-mRuby3). In vivo single-cell electroporation of a Cre-expressing plasmid in *Pdzd8*$^{F/F}$ mice eliminated *Pdzd8* mRNA within 3 days (fig. S1, C to F) and expression of Cre recombinase in cortical neurons reduced PDZD8 protein levels within 7 days (fig. S1, G and H). We performed all imaging experiments no less than 7 days after electroporation. Our acute, single-cell genetic deletion in adult CA1PNs circumvented potential non–cell-autonomous (*22*, *35*) and developmental (*16*) effects that might have arisen from broad and constitutive perturbations. During head-fixed spatial navigation on a cue-rich, 2-m treadmill belt (*36*, *37*), we simultaneously monitored activity dynamics in one dendritic focal plane, which ranged from basal dendrites in stratum oriens to distal tuft dendrites in stratum lacunosum moleculare (fig. S2), and in a somatic plane that often contained coplanar segments of basal dendrite (Fig. 1, B and C, and Movie 1). For motion correction (fig. S3) and dendritic

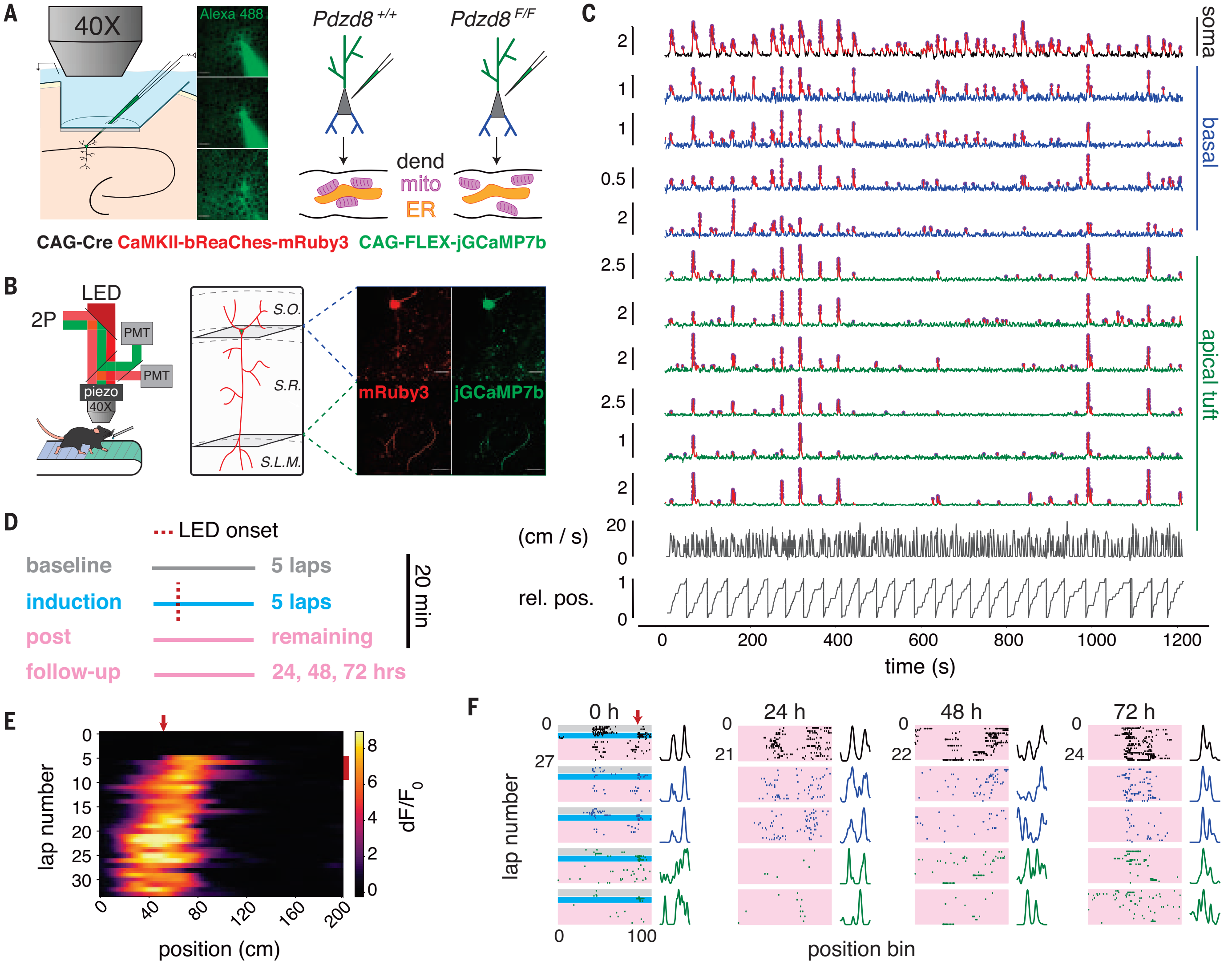


Fig. 1. Simultaneous imaging of somatic and dendritic activity in single, spatially tuned CA1PNs with normal and augmented ICR. (**A**) Left: An electroporation pipette containing plasmid DNAs (listed below) is guided through a silicon-protected slit in an imaging window implanted over dorsal CA1. Middle: Two-photon images before (top), during (middle), and after (bottom) electroporation of plasmid solution to a single cell in the pyramidal layer. Scale bar, 20 μm. Right: Plasmid mixture is delivered through single-cell electroporation to CA1PNs in *Pdzd8*$^{+/+}$ and *Pdzd8*$^{F/F}$ mice. Schematic shows Cre-mediated ablation of ER-mitochondria (mito) contacts in dendrite (dend) of *Pdzd8*$^{F/F}$ CA1PN, augmenting cytosolic impact of ICR. (**B**) Left: Multiplane, dual-channel imaging during head-fixed spatial navigation. LED (620 nm) was used for photoactivation of the opsin-expressing cell. Right: Diagram approximating focal planes imaged for an example cell with corresponding motion-corrected, time-averaged fields of view from each channel showing soma/coplanar basal dendrites (top row) and distal apical tuft dendrites (bottom row). Scale bar, 50 μm. Hippocampal strata are indicated by dashed lines. S.O., stratum oriens; S.R., stratum radiatum; S.L.M., stratum lacunosum moleculare. (**C**) Traces of soma (black), as well as basal (blue) and apical tuft (green) dendrites coacquired from example cell in (B). Vertical scale bars indicate dF/F_0. Detected Ca^{2+} transients are shown in red with deconvolved events as magenta circles. Frame-synchronized animal running speed and relative position (rel. pos.) are shown below traces. (**D**) Schematic describing optogenetic PF induction paradigm. (**E**) Somatic activity heatmap from induction session. LED location is denoted by the red arrow at top; LED laps by red bar at right. (**F**) Event rasters and spatial TCs from a single CA1PN showing somatic and subset of dendritic dynamics on induction day (0 hours) and at the 24-, 48-, and 72-hour time points. ROI types are colored as in (C); conditions are colored as in (D). TCs are scaled to maximum values to aid visualization.

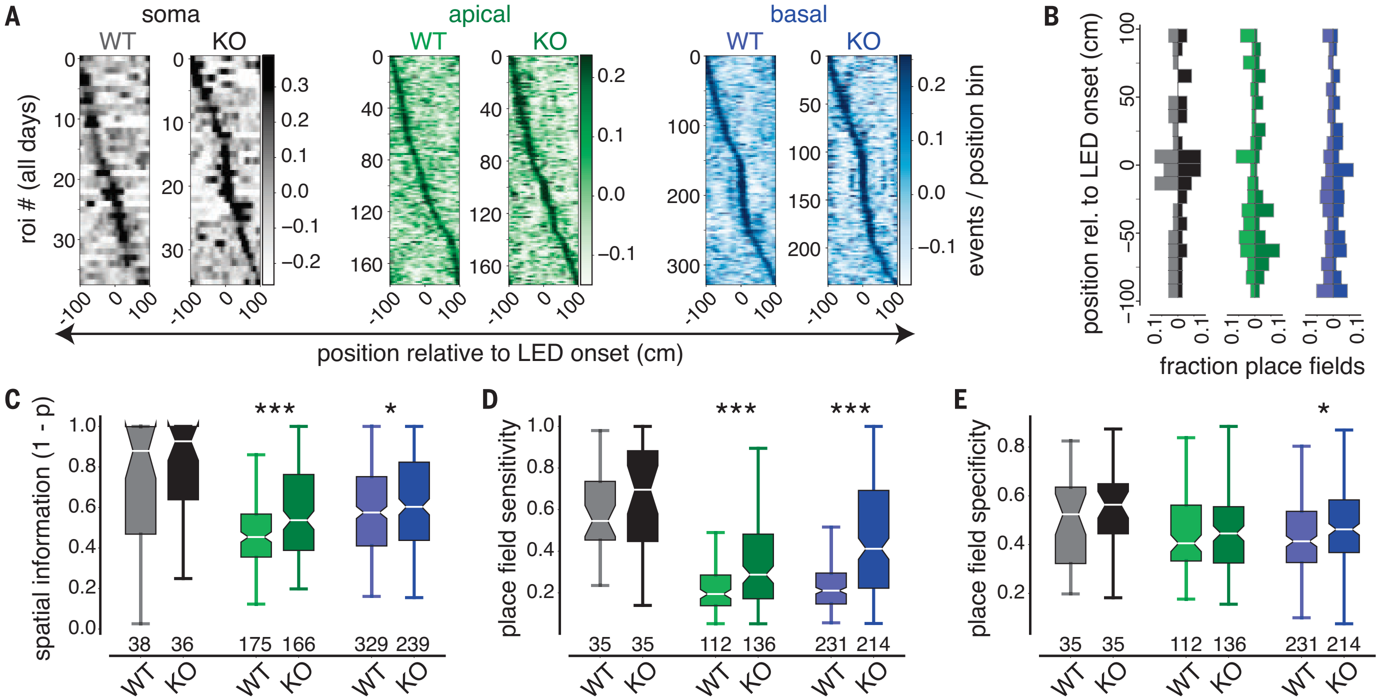


Fig. 2. Effect of augmented ICR on somatic and dendritic spatial tuning properties in single CA1PNs. (**A**) Spatial activity heatmaps showing LED-centered spatial TCs for all *Pdzd8* KO and WT somatic (greys), apical (greens), and basal (blues) ROIs imaged across all days, sorted by TC peak location. (**B**) Mirrored vertical histograms showing distribution of PF center locations relative to LED onset for *Pdzd8* WT (transparent; left of *x*-axis) and KO (opaque; right of *x*-axis) for somatic, apical, and basal ROIs color coded as in (A). (**C** to **E**) PF metrics, including spatial information (C), PF sensitivity (D), and PF specificity (E), shown for all ROI type–genotype combinations. Boxes range from lower to upper quartiles with line at median; whiskers show range of data within 1.5 * (Q3 – Q1) of box boundaries. For (A) to (C), *N* represents the total number of ROIs imaged across days (see table S1 and and the supplementary materials and methods for details). For (D) and (E), *N* represents the total number of ROIs with significant spatial tuning across days. Two-sided unpaired *t* tests and Mann-Whitney *U* tests were used to determine statistical significance. $*P < 0.05$; $***P < 0.001$.

region of interest (ROI) registration (see the supplementary materials and methods), we coacquired static mRuby3 signals through a dedicated collection channel (Fig. 1B).

To evaluate the role of ICR in plasticity and learning, we sought to image CA1PNs expressing PFs. To boost the fraction (*36*, *38*) of CA1PNs functioning as "place cells" and to track the stability of a subset of PFs with a defined time zero, we performed an optogenetic PF induction protocol in each imaged cell (Fig. 1, D and E, and materials and methods). Optogenetically induced PFs displayed characteristic hallmarks (fig. S4, A to D) of naturally occurring in vivo place cell formation (*32*, *33*, *39*, *40*) and shared similar properties with spontaneous, noninduced PFs (fig. S4, E to G). We therefore pooled spontaneous and induced PFs except in longitudinal analyses addressing the stability of spatial feature selectivity, for which a defined time point for PF formation was necessary. We imaged somatic and dendritic activity in 20-min sessions on 4 consecutive days (Fig. 1F and fig. S5), including the initial PF induction day, comparing postinduction dynamics in CA1PNs with normal and augmented ICR.

Acute *Pdzd8* deletion as a tool to investigate the role of ICR

Given the absence of in vivo functional imaging data for apical CA1PN dendrites and the paucity of such data for basal dendrites (*41*–*43*), we first characterized spontaneous Ca^{2+} transient properties in apical and basal dendrites with respect to their parent soma and assessed any impact of our ICR-augmenting manipulation on basic CA1PN activity dynamics. *Pdzd8* KO and WT CA1PN cell bodies did not differ in Ca^{2+} transient frequency or amplitude (fig. S6, A and B). Isolated dendritic transients (no coincident somatic event; see fig. S6, C and D, for examples) were overall rare compared with somatic transients with minor but significant differences in event frequency between compartment and genotype (fig. S6E). Isolated apical dendritic Ca^{2+} transients, which have not previously been reported in CA1PNs, were larger and more frequent than those observed in basal dendrites (fig. S6, E and F). Increasing ICR through *Pdzd8* KO did not affect isolated transient amplitude in either dendritic compartment (fig. S6F). We cannot formally rule out the possibility that undetected somatic action potentials back-propagate into imaged dendrites and contribute to apparent isolated events. However, a difference in event amplitude between genotype would be reasonably expected if ICR were routinely activated in the setting of local dendritic Ca^{2+} transients.

To further assess the potential effects of acute *Pdzd8* KO on baseline Ca^{2+} handling and excitability in adult CA1PNs, we first compared isolated, single-spine Ca^{2+} transients (no coincident shaft-level event; see fig. S6, G and H, for examples) in a separate set of CA1PNs electroporated with pCAG-Cre, pCAG-NES-jRGECO1a, and pCAG-FLEX-eGFP. We imaged individual apical and basal dendritic segments at high magnification in $Pdzd8^{F/F}$ and WT control mice under light anesthesia on a treadmill belt lacking spatial cues (fig. S6, G and H, and supplementary materials and methods). Although isolated spine Ca^{2+} transients were generally of greater amplitude and frequency in apical versus basal dendrites, these event properties did not significantly differ between *Pdzd8* KO and WT CA1PNs (Fig. S6, I and J). Next, we compared the electrophysiological properties of *Pdzd8* KO and WT CA1PNs after targeting either AAV-CaMKII-Cre-mCherry or mCherry-only control virus to dorsal CA1 of

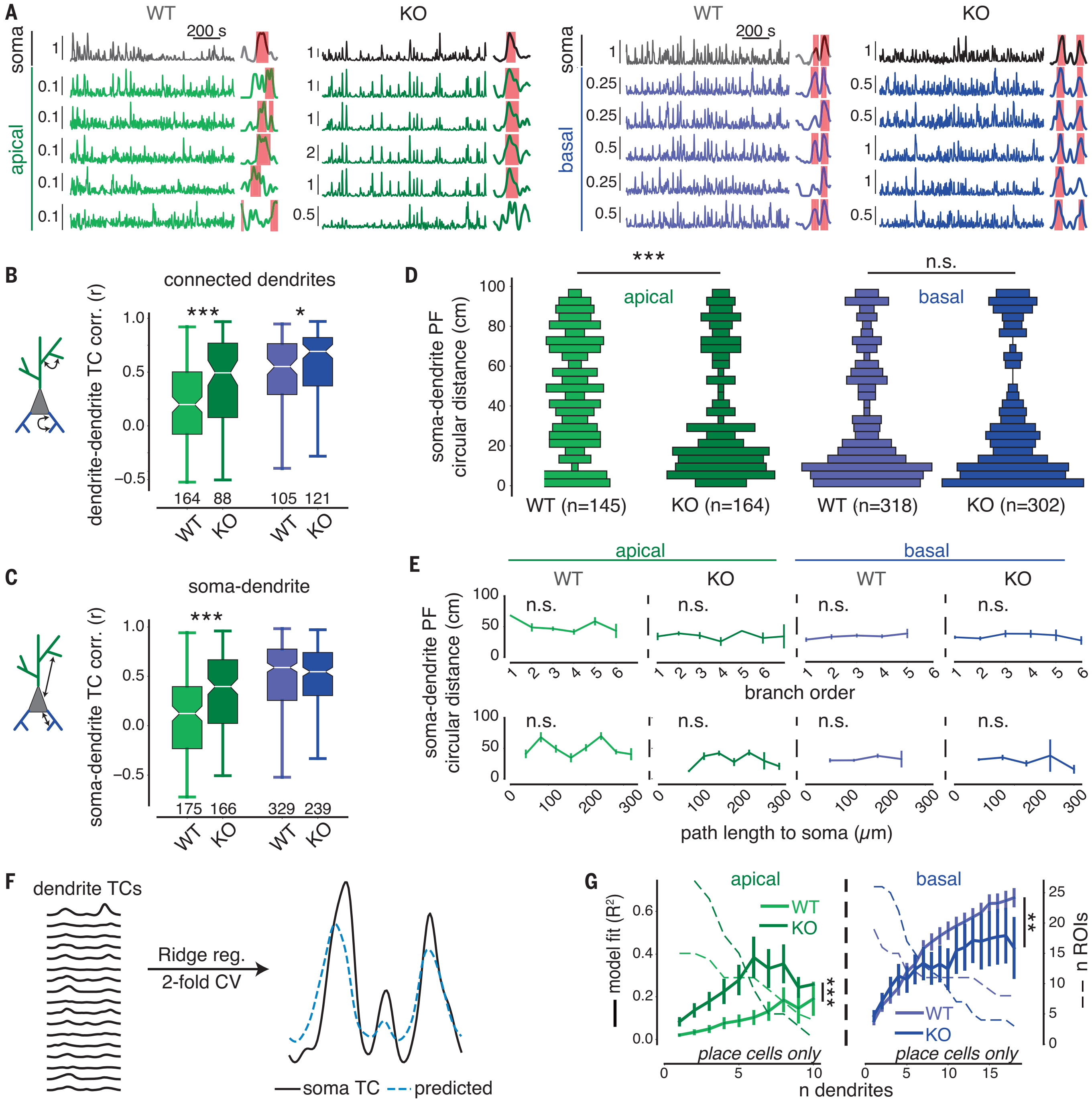


Fig. 3. Compartment-specific regulation of dendrite-dendrite and soma-dendrite co-tuning by ICR. (**A**) Example dF/F_0 traces and TCs for all ROI type-genotype combinations. Vertical scale bars indicate dF/F_0. Dendritic traces and TCs are grouped under corresponding somata. Detected PFs are denoted by pink shaded areas overlaid onto TCs. (**B** and **C**) TC correlations calculated exclusively between connected, co-imaged dendrites (B) (see fig. S8) and between dendritic ROIs and their corresponding somata (C). *N* represents the total number of ROIs imaged across days (table S1). Boxes range from lower to upper quartiles with line at median; whiskers show range of data within 1.5 * (Q3 − Q1) of box boundaries. (**D**) Vertical histograms showing distributions of minimum circular distances between somatic and dendritic PF centers. Bar widths represent relative abundance of bin values. *N* represents the total number of PFs from imaged dendrites belonging to place cells across days. (**E**) Circular distances in (D) plotted against branch order (top) and path length to soma binned every 40 μm (bottom). Pearson correlation analysis was used on nonbinned data. (**F**) Schematic for predicting somatic TC based on TCs of connected dendrites (see the supplementary materials and methods) with example performance shown at right. (**G**) Model performance (R^2) plotted against the number of dendrites included in training data. Number of cell sessions containing *N* dendrites is shown in dashed lines on the second *y*-axis. Note that measurement precision decreases with *N*. Apical two-way ANOVA, genotype effect: $F_{(1,43)} = 37.39$, $P < 0.001$); *n* dendrites effect: $F_{(9,217)} = 4.57$, $P < 0.001$; interaction: $F_{(9,217)} = 0.81$, $P > 0.05$; *n* = 16 WT and 28 KO place cells. Basal two-way ANOVA, genotype effect: $F_{(1,44)} = 7.35$, $P < 0.01$; *n* dendrites effect: $F_{(17,387)} = 13.44$, $P < 0.001$; interaction: $F_{(17,387)} = 1.05$, $P > 0.05$; *n* = 19 WT and 26 KO place cells. Error bars represent SEM. Distributions in (B) to (D) were compared using two-sided unpaired *t* tests and Mann-Whitney *U* tests. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

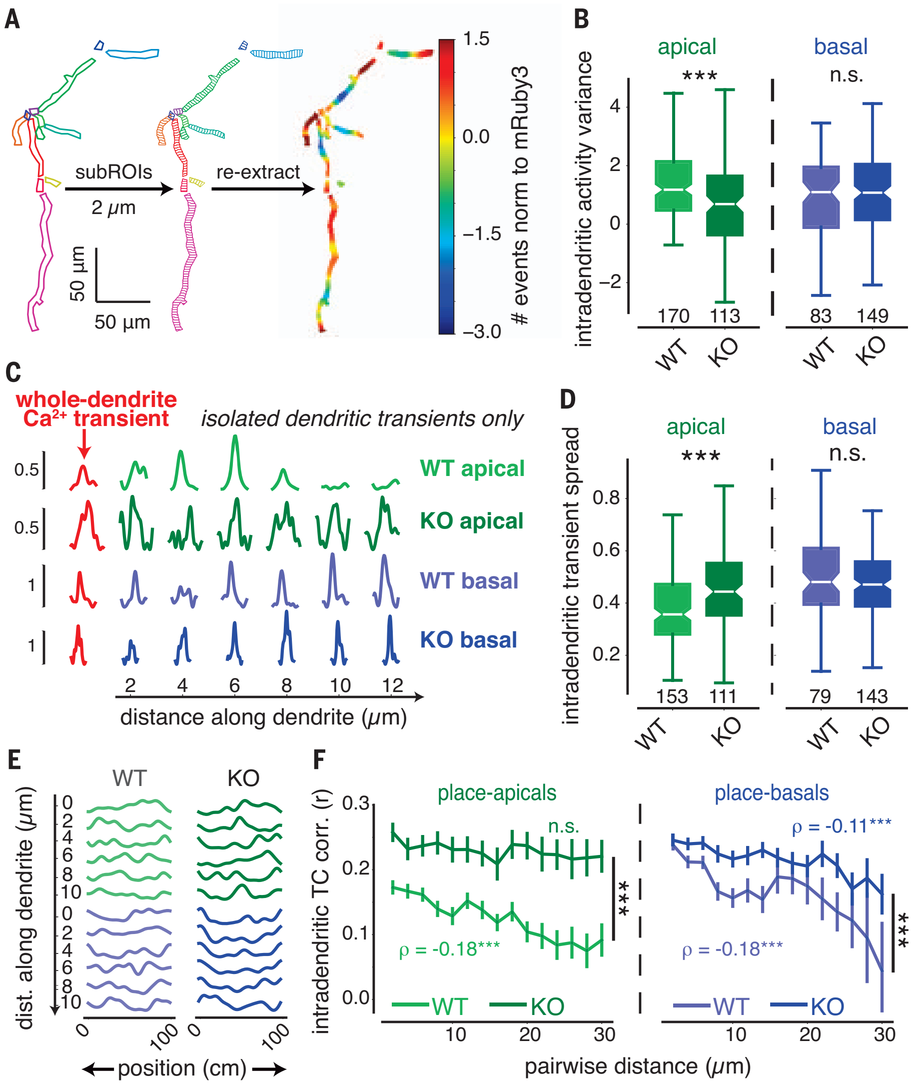


Fig. 4. Intracellular Ca^{2+} release shapes activity dynamics and spatial feature selectivity within single dendritic branches of CA1 PNs. (**A**) Approach to analyzing intradendritic dynamics. Each ROI was segmented into 2-μm sub-ROIs along its longest axis. Signals were then re-extracted from each sub-ROI for further analysis. Heatmap shows activity distribution for each imaged apical dendrite from a single WT CA1PN. Activity was normalized within ROIs to static mRuby3 signal intensity ($\log_2$ ratio) to control for differences in focality (see the supplementary materials and methods). (**B**) Within-dendrite variance of activity levels between WT and *Pdzd8* KO apical and basal dendrites ($\log_2$ ratio of coefficients of variation for events and mRuby3 signal intensity). (**C**) Spatial spread of isolated Ca^{2+} transients within single dendrites. For each Ca^{2+} transient detected from a whole-dendrite ROI (red traces), corresponding dF/F_0 signals are plotted from the first six sub-ROIs (12 μm) of that dendrite. (**D**) Normalized intradendritic spread of isolated Ca^{2+} transients within WT and *Pdzd8* KO apical and basal dendrites. A value of 1.0 indicates uniform sub-ROI peak amplitudes (see the supplementary materials and methods). (**E**) Sub-ROI spatial TCs, color coded as in (C). (**F**) Intradendritic TC correlations plotted as a function of distance separating two sub-ROIs (see the supplementary materials and methods). Spearman correlation coefficients are shown on plots. Distances were compared between genotypes using Mann-Whitney *U* tests. *N* = 110 WT and 100 KO place-apicals; 223 WT and 184 KO place-basals. Error bars represent SEM. Boxes range from lower to upper quartiles with line at median; whiskers show range of data within 1.5 * (Q3 – Q1) of box boundaries. Distributions in (B) and (D) were compared using two-sided unpaired *t* tests. ***$P < 0.001$.

adult *Pdzd8*$^{F/F}$ mice (fig. S7A and supplementary materials and methods). Apart from minor but significant differences in action potential amplitude and half-width, *Pdzd8* KO and WT CA1PNs shared similar electrophysiological attributes (fig. S7, B to K).

ICR controls the diversity and reliability of dendritic spatial tuning in single CA1PNs

We reasoned that if ICR regulates synaptic plasticity in vivo, then augmenting its spread and magnitude should affect specific aspects of dendritic feature selectivity in CA1PNs. These effects would in turn provide insight into the endogenous function of ICR. We thus surveyed spatial tuning properties of apical and basal dendrites with respect to their cognate soma in individual *Pdzd8* KO and WT CA1PNs. Because optogenetically induced PFs represented only a fraction of those observed across all 4 days (Fig. 2, A and B), and induced PFs shared similar characteristics with spontaneous PFs (fig. S4, E to G), we initially analyzed PFs regardless of their location relative to prior light-emitting diode (LED) stimulation. In *Pdzd8* KO cells with augmented ICR, both apical and basal dendritic activity showed increased overall spatial tuning (Fig. 2C and supplementary materials and methods). *Pdzd8* KO apical and basal "place dendrites," i.e., dendrites in which PFs were detected, fired more reliably upon the animal's traversal of their PFs (Fig. 2D), whereas basal place dendrites fired more selectively within their PFs (Fig. 2E). Increasing the cytosolic impact of ICR did not significantly affect any measure of somatic spatial tuning (Fig. 2, C to E). This is consistent with the previous finding that increased within-field basal dendritic activity does not portend increased CA1PN somatic activity (*41*) but may also be due to relatively low sample sizes for somatic ROIs; all three metrics in Fig. 2, C to E, trended higher for *Pdzd8* KO CA1PN somata. In conclusion, although augmenting the cytosolic impact of ICR did not affect measures of baseline Ca^{2+} handling or excitability (figs. S6 and S7), the same manipulation promoted feature selectivity in CA1PN apical and basal dendrites.

We next tested whether ICR influences the distribution of dendritic spatial tuning preferences (Fig. 3A). Augmenting ICR by *Pdzd8* KO significantly increased the correlation of spatial tuning curves (TCs) among connected apical and basal dendritic branches (Fig. 3B; see fig. S8 for an illustration of the analytic approach) as well as between apical dendrites and their soma (Fig. 3C). Soma-dendrite TC correlations were unaffected in basal dendrites, although WT basal dendrites already displayed high dendrite-dendrite and dendrite-soma correlations that may have led to a ceiling effect (Fig. 3, B and C). To exclude dendrites that exhibited minimal spatial tuning to begin with, we next restricted our analysis to place dendrites that belonged to place cells and

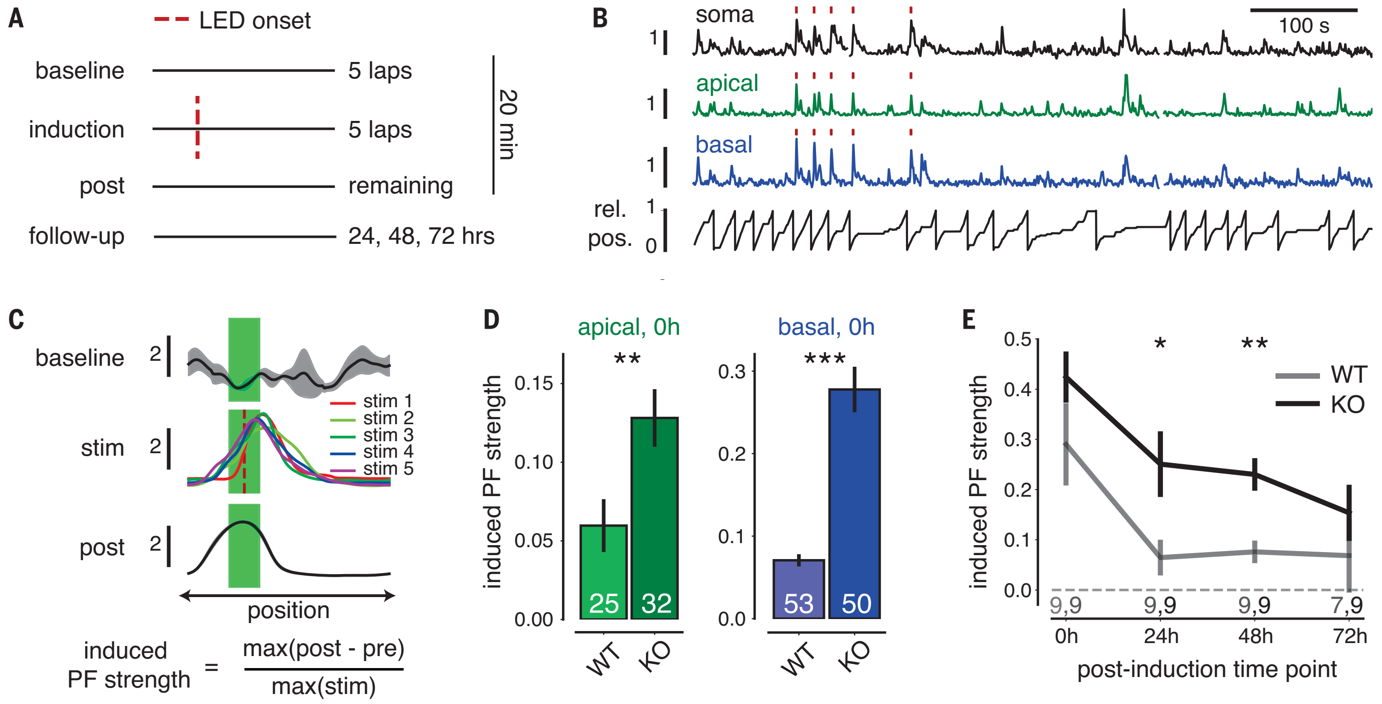


Fig. 5. Dendritic strength and somatic stability of optogenetically induced PFs in single WT and *Pdzd8* KO CA1PNs. (**A**) Paradigm for optogenetic PF induction as in Fig. 1D. (**B**) Traces showing LED-evoked and naturally occurring activity in simultaneously acquired somatic, apical, and basal jGCaMP7b traces from a single CA1PN with relative animal position (rel. pos.) plotted below. LED stimulation during induction laps indicated by red ticks. (**C**) Quantifying strength of induced spatial tuning. Top: Mean somatic dF/F_0 by position from baseline, induction (stim), and post laps for an induced cell with baseline and postactivity shown as mean ± SEM and induction laps shown individually. Red dashed line indicates LED onset and green shaded area denotes "LED zone" used for quantifying induced activity. Bottom: Quantification of induced PF strength. Signals are normalized to maximum LED response to control for cell-to-cell variability in jGCaMP7b expression level, excitability, and/or focality. (**D**) Apical and basal induced PF strength for successfully induced dendrites of WT and *Pdzd8* KO CA1PNs on day 0. Two-sided unpaired *t* tests were used to determine statistical significance. (**E**) Somatic induced PF strength of successfully induced CA1PNs across days. Mixed-effects model with post hoc *t* tests, genotype effect: $F_{(1,16)} = 7.80$, $P < 0.05$; time effect: $F_{(3,46)} = 10.85$, $P < 0.001$; interaction: $F_{(3,46)} = 0.32$, $P > 0.05$. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. Error bars and shaded error bands indicate SEM.

calculated the minimum circular distances along the cued belt that separated each dendritic PF from that of its soma. This analysis revealed a previously unappreciated degree of heterogeneity in spatial tuning preference among WT apical dendrites relative to their somata and, by contrast, a clear tendency toward somatic co-tuning in WT basal dendrites (Fig. 3D). Moreover, increased ICR was associated with a significant shift in apical dendritic PFs toward those of their soma; *Pdzd8* KO apical dendrites were as co-tuned to their soma as were WT basal dendrites (Fig. 3D). Consistent with the observation that *Pdzd8* KO selectively affected apical dendrite TC correlations (Fig. 3, B and C), *Pdzd8* KO did not affect basal dendrite-soma co-tuning (Fig. 3D). Dendritic PF specificity did not attenuate with increasing soma-dendrite PF distance (fig. S9) and apical PFs in *Pdzd8* KO CA1PNs displayed increasing specificity the more anti-tuned they were with their somatic PFs (fig. S9A, right). Additionally, neither branch order nor path length predicted dendrite-soma co-tuning (Fig. 3E). Finally, we formally tested the extent of dendrite-soma co-tuning by training a regularized linear model to predict somatic TCs based on those of connected dendrites (Fig. 3F and supplementary materials and methods). This model revealed that apical dendrites with augmented ICR markedly outperformed WT apical dendrites in predicting their somatic TCs, whereas the same manipulation caused basal dendrites to marginally underperform (Fig. 3G).

ICR differentially shapes the subcellular distribution of spatial tuning within single apical and basal dendritic segments

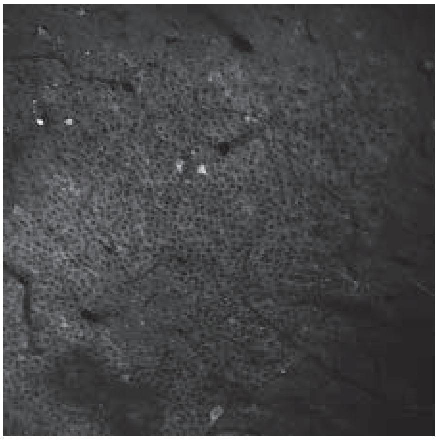
Our results show that the amount of Ca^{2+} released from the ER into the cytosol influences CA1PN dendritic spatial tuning in vivo, but how does the extent of ICR fit into plasticity mechanisms underlying hippocampal feature selectivity? The presence of ER in dendritic spines positively predicts spine head size (*8*), suggesting a role for ICR in spine structural plasticity. In vitro experiments support such a role (*44*–*46*) and further suggest that the spread of ICR along a dendritic branch in part determines which spines will undergo synaptic plasticity (*47*, *48*). Therefore, the spatial extent of ICR along a dendritic branch may ultimately influence local integrative properties underlying dendritic spatial receptive fields in vivo. We investigated how increasing the cytosolic impact of ICR would influence activity and feature selectivity within single dendrites, i.e., across nearby dendritic spines. We subdivided dendritic ROIs into 2-μm segments ("sub-ROIs") and re-extracted, processed, and analyzed sub-ROI signals (Fig. 4A). To account for within-dendrite differences in focality, we normalized activity measures to static mRuby3 signal intensity (see the supplementary materials and methods). Variability in overall activity levels between sub-ROIs was selectively reduced in *Pdzd8* KO apical dendrites (Fig. 4B). Next, we investigated whether increasing ICR altered the spatial features of isolated Ca^{2+} transients along individual dendritic segments (Fig. 3C; see fig. S10A for additional examples). Indeed, genetic amplification of ICR increased the spread of isolated Ca^{2+} transients specifically in apical dendrites (Fig. 4D; fig. S10, B to E; and supplementary materials and methods).

Finally, we calculated TC correlations between all possible combinations of sub-ROIs and investigated how co-tuning varied between sub-ROIs as a function of the anatomical distance separating them along the dendritic arbor (Fig. 4E). Both apical and basal WT dendrites showed a monotonic decrease in intradendritic TC correlation with distance within single dendritic branches (Fig. 4F). It has been previously suggested (*16*, *49*–*52*), and recently demonstrated (*53*), that feature-correlated inputs are spatially clustered along individual CA1PN dendrites. The distance-dependent drop-off in TC correlations that we observed in dendrites of WT CA1PNs corresponds to the reported distance along which presynaptic inputs were reported to cluster: ~10 μm (*53*). Given the relatively low sensitivity of genetically encoded Ca^{2+} indicators (GECIs), such as jGCaMP7b (*54*), we took this correspondence to indicate that intradendritic GECI signals can provide meaningful insight into how spatially modulated presynaptic inputs are integrated postsynaptically. Similarly, by analyzing spatial tuning within the dendrites of *Pdzd8* KO CA1PNs, we found that increasing ICR strengthened intradendritic TC correlations in both compartments and erased the distance-dependent drop-off in TC correlations specifically within apical dendrites (Fig. 4F).

ICR strengthens dendritic feature selectivity and stabilizes output-level receptive fields

CA1 place cells can gain, lose, remap, or retain their spatial tuning properties. Exposure to a new environment promotes PF formation and remapping (*39*, *55*) whereas, upon repeated exposure to the same environment, subsets of cells stably represent specific locations (*38*, *56*, *57*). Given the role of ICR in controlling the spatial extent of feature selectivity within and across dendrites and in promoting reliability in dendritic PFs, we tested how these input-level PF properties ultimately influence output-level, i.e., somatic, spatial tuning over days. To assess the relative stability of CA1PN spatial tuning over days, we restricted our analysis to activity dynamics within a nominal "LED zone" centered around the location of LED onset during our optogenetic induction protocol (Fig. 5, A to C). We longitudinally tracked somatic activity in the same environment across 72 hours and assessed dendritic activity specifically on day 0 (see the materials and methods). Because PF formation requires both pre- and postsynaptic activity, and only a subset of synapses should receive excitatory presynaptic input at the time of postsynaptic optogenetic stimulation, we expected to induce a fraction of dendrites. Still, dendrites of CA1PNs with augmented ICR were more likely to display optogenetically induced spatial tuning (fig. S11A), revealing a role for ICR in the establishment of dendritic feature selectivity. To measure the strength of induced spatial tuning, we focused specifically on ROIs that were successfully induced (see the materials and methods). We first accounted for cell-to-cell differences in plasmid expression, excitability, and/or focality by normalizing dF/F_0 from each ROI to its peak LED response. We then quantified induced PF strength based on postinduction activity near the location of LED onset relative to a baseline measurement (Fig. 5C).

Both apical and basal dendrites of *Pdzd8* KO CA1PNs displayed stronger induced PFs on day 0 (induction day; Fig. 5D). Although *Pdzd8* KO cell bodies did not show significantly stronger PFs on day 0 relative to WT cells, consistent with our observation that increased dendritic spatial tuning did not significantly translate to output-level PFs (Fig. 2, C to E), induced somatic PFs were more stable over days (Fig. 5E). We considered the possibility that backpropagating action potentials (bAPs) might explain the appearance of strong dendritic PFs in *Pdzd8* KO CA1PNs. If the soma of a *Pdzd8* KO CA1PN fires more robustly near the LED zone, then somatic action potentials may more effectively invade the dendritic arbor and drive increased voltage-dependent Ca^{2+} influx from the extracellular space (*58*). However, dendritic PFs were stronger in *Pdzd8* KO CA1PNs on day 0, when somatic PF strengths were similar (Fig. 5, D and E). Additionally, dendritic proximity to soma did not predict soma-dendrite co-tuning (Fig. 3E). Therefore, we conclude that induced dendritic PF strength is not contaminated by bAPs in our measurements and reflects the degree of local spatial tuning.



Movie 2. Population GCaMP6f dynamics in the CA1 pyramidal layer during random foraging spatial navigation as in Fig. 6. Frames are motion corrected and shown at actual speed. Data were acquired using two-photon microscopy with a 16× objective lens at 1.2× optical zoom.

ICR shapes PF formation and stabilizes population-level spatial representations

To more rigorously assess the impact of increasing ICR on output-level feature selectivity, and to do so in the context of naturally occurring spatial tuning, we imaged large, intermingled populations of *Pdzd8* KO and WT CA1PN cell bodies (Movie 2 and fig. S12D; see the materials and methods for classification strategy) as mice navigated a familiar cued treadmill belt over 5 consecutive days (Fig. 6, A to C). ROIs were registered across days for longitudinal monitoring of activity dynamics and spatial tuning (Fig. 6C, bottom, and fig. S12). Across two signal extraction and neuropil decontamination approaches, *Pdzd8* KO and WT CA1PNs did not consistently differ in basic Ca^{2+} transient properties or baseline GCaMP6f fluorescence levels (fig. S13). This large-scale, population-level analysis uncovered minor but significant changes in somatic PF properties (fig. S14, A to D) that were not detectable with soma sample sizes in single-cell experiments (Fig. 2, C to E). Consistent with optogenetic PF induction experiments (Fig. 5), somatic TCs of *Pdzd8* KO place cells were more stable day to day relative to WT control place cells (fig. S14G).

One limitation of our single-cell imaging experiments is that their cross-subject design did not permit comparison of PF properties that closely correlate with animal behavior. PF width correlated with periformation velocity (fig. S4C), as previously reported (*33*), and the backward shift in PF peak activity (fig. S4D) that characterizes BTSP-mediated PF formation (*33*) covaried with these measures (fig. S11, B to E). Because imaging *Pdzd8* KO and WT CA1PNs within animal controlled for potential behavior-related biases, we sought to identify and compare naturally occurring PF formation events from our mixed-population data. To do this, we devised an automated PF formation detection approach specifically tailored to BTSP-like events (Fig. 6D, fig. S15A, and materials and methods). Augmenting ICR through *Pdzd8* KO did not affect the rate of PF formation across days (Fig. 6E). By assessing place cell TCs relative to the date of PF formation, we again observed that *Pdzd8* KO stabilized spatial TCs of newly formed PFs over days (Fig. 6F). We next assessed PF peak shifts by aligning postformation TCs to the location of PF formation (Fig. 6G and materials and methods). In contrast to WT control CA1PNs, which displayed the backward shift previously established (*33*) and observed in single-cell optogenetic PF induction experiments (fig. S4D and fig. S11E), *Pdzd8* KO CA1PNs with augmented ICR showed no discernable postformation shift on average (Fig. 6H; see fig. S15B for additional examples). Finally, whereas PF width correlated with periformation velocity in both populations, *Pdzd8* KO CA1PN PF widths increased more sharply with velocity and exceeded those of WT control CA1PNs on the whole (Fig. 6, I and J).

Discussion

We found that release of Ca^{2+} from the ER plays a key role in synaptic plasticity mechanisms underlying dendritic receptive field formation, regulating feature selectivity at multiple levels of organization from intra- and interdendritic levels to dendrite-soma co-tuning. These dendritic functions coalesce to govern multiple properties of output-level spatial tuning. The combination of approaches that we developed offers new insights into how ICR participates in PF formation. Based on the impact of increased ICR on two PF characteristics that are specifically influenced by animal behavior during plasticity induction, PF width and peak shift, we conclude that ICR plays an active role in the emergence of cellular feature selectivity in the hippocampus. Consistent with its dependence on voltage-gated Ca^{2+} entry (59), ICR appears to operate downstream of plateau potential initiation: Somatic PF stability was improved both in spontaneously occurring and in optogenetically induced PFs, in which plateau potentials and their associated global Ca^{2+} influx (32, 60) are obviated by direct optogenetic depolarization.

It may seem counterintuitive that the spread of ICR would be important if plateau potentials (or optogenetic stimulation) already involve global Ca^{2+} influx from the extracellular space. However, voltage-gated influx of extracellular Ca^{2+} associated even with high

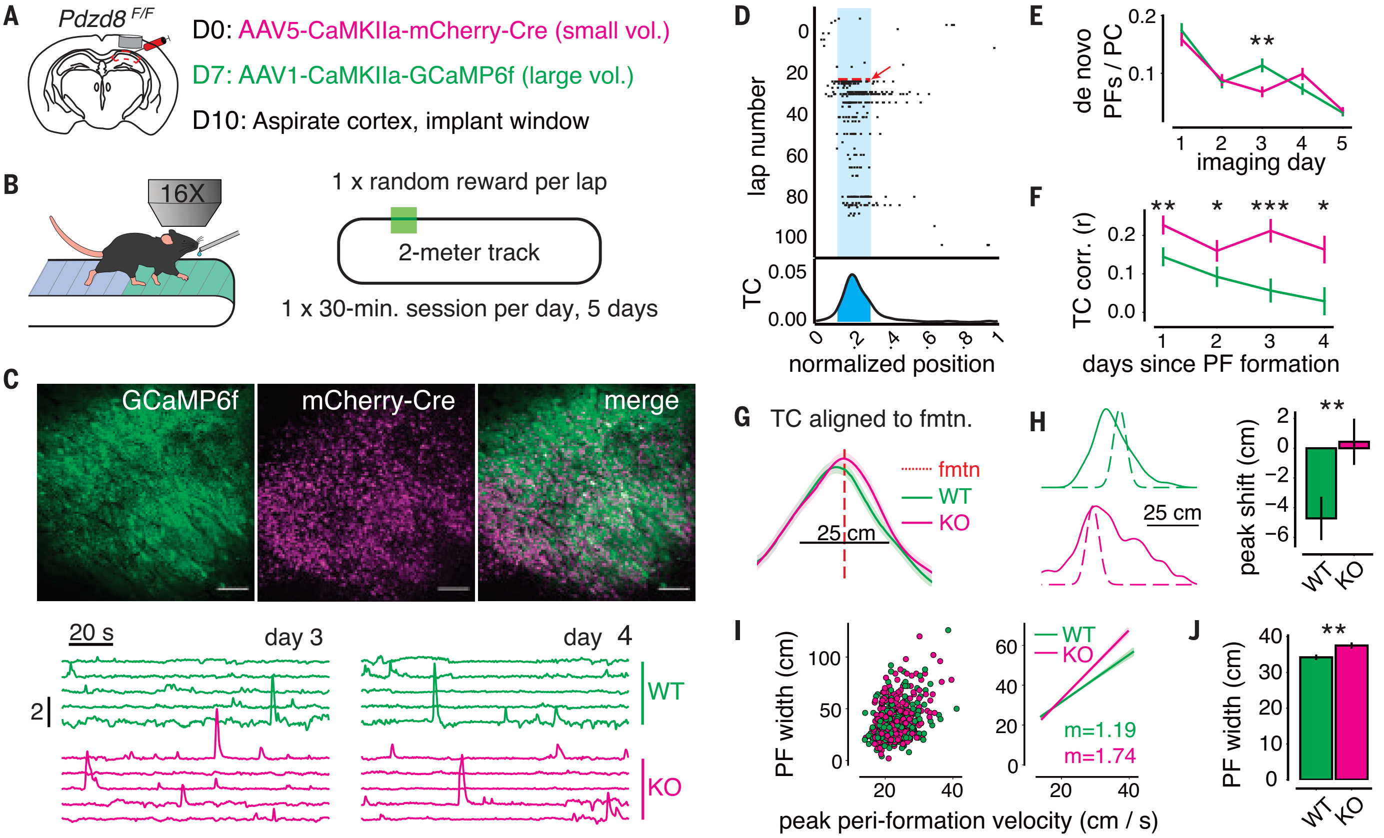


Fig. 6. Large-scale, mixed-population imaging of WT and *Pdzd8* KO CA1PNs implicates intracellular Ca^{2+} release in the formation and stability of output-level feature selectivity. (**A**) Viral strategy. AAV encoding mCherry-Cre fusion protein was injected days before GECI-encoding AAV at low volume to allow for recombination/protein clearance and robust mCherry signal. (**B**) Signals were acquired at 16× (1.2 to 1.5× optical zoom) during a head-fixed spatial navigation paradigm with randomly located reward. Identical fields of view (FoVs) were imaged across 5 consecutive days (see fig. S12). N = 2719 WT and 3260 KO cells from six mice. (**C**) Top: FoVs showing GCaMP6f (left) and mCherry-Cre (middle) expression with overlay (right). Scale bar, 100 μm. Bottom: GCaMP6f traces from the same *Pdzd8* KO (magenta) and WT (green) ROIs imaged on separate days. Vertical scale bar represents dF/F_0. (**D**) Plot showing automated detection of de novo PF formation event. Top: Deconvolved events (black ticks) plotted by lap number and position along cued belt. Red dashed line and arrow indicate detected formation lap with PF bounds shaded in cyan. Bottom: TC for above raster plot with PF shaded in cyan. (**E**) Rates of de novo PF formation over days for all place cells. Two-way ANOVA with post hoc t tests, genotype effect: $F_{(1,1808)} = 0.72$, $P > 0.05$; time effect: $F_{(4,7806)} = 46.88$, $P < 0.001$; interaction: $F_{(4,7806)} = 3.16$, $P < 0.05$. (**F**) TC correlations (TC corrs) over days from *Pdzd8* KO and WT CA1PNs that displayed de novo PF formation events relative to the day of formation. Two-way ANOVA with post hoc t tests, genotype effect: $F_{(1,673)} = 24.78$, $P < 0.001$; time effect: $F_{(3,2016)} = 3.36$, $P < 0.05$; interaction: $F_{(3,2016)} = 0.99$, $P > 0.05$. (**G**) Mean, postformation TCs from all detected de novo events, aligned to where PFs were initially formed (see the supplementary materials and methods). (**H**) Left: Example *Pdzd8* WT (green) and KO (magenta) postformation TCs (solid line) overlaid onto TC from formation lap (dashed line). Right: Quantification of peak shift, i.e., distance between peaks of formation and postformation TCs. N = 350 and 369 formation events from *Pdzd8* WT and KO CA1PNs, respectively. CA1PNs with multiple PFs were excluded from analysis. (**I**) Relationship between periformation velocity and resultant PF width for *Pdzd8* WT (green, N = 366) and KO (magenta, N = 373) events across six mice. Left: Scatter plot of all events. Right: Least-squares linear fits of data shown on left with corresponding slopes (m) ($r_{WT} = 0.33$, $P < 001$; $r_{KO} = 0.41$, $P < 0.001$). (**J**) Widths of all de novo PFs. Error bars represent SEM. Boxes range from lower to upper quartiles with line at median; whiskers show range of data within 1.5 * (Q3 − Q1) of box boundaries. Distributions were compared using two-sided unpaired t tests and Mann-Whitney U tests. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

degrees of spiking exerts a minor impact on intracellular Ca^{2+} concentrations ($[Ca^{2+}]_i$) relative to ICR (10, 14). Consistent with the high threshold for ICR engagement (5, 10–13, 59), and thus its presumed rarity in vivo, *Pdzd8* deletion did not alter basic excitability or Ca^{2+}-handling properties that might otherwise have obfuscated interpretation of GECI signals. Therefore, altered ICR alone is unlikely to account for our observations using GCaMP to monitor voltage-dependent influx of extracellular Ca^{2+}. Rather, we propose that ICR occurs relatively infrequently in vivo—during plateau potentials in CA1PNs, for example—and participates in the induction of synaptic plasticity. In this context, our results suggest that augmented ICR lastingly modifies dendritic integrative properties.

The widening of *Pdzd8* KO PFs specifically in the forward direction suggests that a prolonged ICR-driven Ca^{2+} signal potentiates more dendritic spines receiving input after plateau potential onset. Consistent with a broadening of potentiated inputs, we found that spatial tuning preferences were more uniform within single apical dendritic segments of *Pdzd8* KO CA1PNs. It was recently established that feature-correlated presynaptic inputs are spatially clustered along CA1PN dendrites, i.e., that neighboring spines tend to receive input corresponding to similar sensory experiences (53). We observed a postsynaptic vestige of this phenomenon in WT dendritic segments: Intradendritic TC correlations systematically decreased with intervening distance. This drop-off began after ~10 μm, the same distance across which correlated presynaptic inputs were previously reported to cluster. The fact that augmenting ICR abrogated this distance-dependent decorrelation in apical dendrites and widened output-level PFs additionally suggests that the diffusion of ER-released Ca^{2+} along a dendritic shaft may influence the number of spines co-potentiated during plasticity induction. Although previous in vitro studies directly support such a model (47, 48), ICR may also regulate intradendritic spatial tuning through branch-specific plasticity of dendritic excitability (61). Given that *Pdzd8* KO did not affect baseline GCaMP fluorescence or the amplitude of isolated events in dendritic spines and shafts, it is unlikely that acute effects of augmented ICR, which again is presumed to be relatively infrequent in vivo, influence intradendritic spatial tuning. In addition to broadening the anatomical distribution of dendritic feature selectivity, increasing ICR also strengthened spatial tuning: Dendritic and somatic PFs were more specific and sensitive, whereas somatic PFs were also more durable over time in the face of constant representational turnover within CA1 (38, 62). Finally, the substantial impact of increasing ICR on dendrite-dendrite and soma-dendrite co-tuning, again specifically within the apical arbor, indicates that ICR operates on multiple spatial scales (12, 13, 19) to shape output-level receptive fields.

A role for ICR in the emergence of feature selectivity is further supported by the velocity-dependent nature in which augmenting ICR widened PFs. CA1PN synapses require temporally correlated pre- and postsynaptic activity to undergo plasticity (33, 60, 63). Higher running speeds positively modulate the rate of presynaptic input onto CA1PNs (64, 65), increasing the pool of synapses eligible for potentiation during PF formation in vivo and hence the velocity-PF width relationship (33). Therefore, whereas unperturbed ICR may reach and potentiate most eligible inputs at lower velocities, increased ICR would be predicted to more comprehensively engage larger pools of eligible connections generated during bouts of increased running speed. We acknowledge that normative statements regarding ICR function are difficult to make based on a gain-of-function manipulation. However, our mild, ~50% augmentation (9) of $[Ca^{2+}]_i$ associated with ICR may well fall within the range of physiological variability, because the magnitudes of individual ICR events are known to vary according to a number of factors (10, 12, 13, 16, 59, 66, 67). Therefore, although we acknowledge that our findings pertain to one specific plasticity mechanism underlying PF formation, we propose ICR magnitude, as regulated by ER-mitochondria contact sites (9), as a general mechanism influencing the distribution of synaptic plasticity induction in vivo.

Independent of ICR, we found that apical dendrites of CA1PNs have a great deal of functional autonomy from the soma. Apical CA1PN dendrites have been the subject of extensive in vitro research due to their presumed central role in hippocampus-dependent learning, but their in vivo activity had remained elusive. The broadly distributed tuning preferences observed across apical dendrites in single WT CA1PNs could not be explained by path length from soma or by branch order, nor did soma-basal dendritic co-tuning track with these morphological properties. This is consistent with a prior observation that co-firing between basal dendrites and their soma does not depend on intervening distance (41) and indicates that somatic AP back-propagation does not appreciably influence soma-dendrite co-tuning, at least across the relatively broad ranges of branch order and path length that we sampled. Nor were anti-tuned dendritic PFs of lesser quality: PF specificity did not track with the degree of soma-dendrite co-tuning. Therefore, similar to pyramidal neurons of various sensory cortices (68–70), CA1PNs are able to integrate diverse dendritic tuning preferences into single, clean receptive fields. How this aspect of feature selectivity emerges remains poorly understood.

By contrast, basal dendrites showed stronger correlation with somatic tuning. This may have been due to a true biological difference between apical and basal dendrites of CA1PNs. Alternatively, somatic action potentials may invade basal dendrites more efficiently because their relatively closer proximity to the soma would mitigate distance-dependent attenuation of bAPs (58). In this case, correlated voltage-dependent Ca^{2+} influx from the extracellular space would indeed be expected. However, given that soma-dendrite co-tuning was predicted by neither branch order nor path length in either dendritic compartment, we propose that apical dendritic tuning is more decoupled from somatic tuning relative to basal dendrites and that ICR, as regulated by mitochondrial Ca^{2+} buffering, contributes to this functional compartmentalization. The compartment-specific nature of ICR's role in plasticity intersects with the CA1PN input structure in a potentially powerful way. Given that apical and basal dendrites receive synaptic input from unique combinations of afferent circuits carrying distinct streams of information (71), compartment-specific action by ER may allow selected features of experience to preferentially influence learning.

Methods Summary

See the supplementary materials for full materials and methods information.

Animals

All experiments were conducted in accordance with National Institutes of Health guidelines and were approved by the Columbia University Institutional Animal Care and Use Committee under protocols AC-AABF6554, AC-AAAX1460, and AC-AAAO0850. Animal health and welfare were supervised by a designated veterinarian. Columbia University animal facilities comply with all appropriate standards of care, including cage conditions, space per animal, temperature, light, humidity, food, and water. Adult (postnatal day 60 or older) male and female $Pdzd8^{F/F}$ and WT mice, fully backcrossed to the C57Bl/6 genetic background, were used for all experiments. The authors are not aware of an influence of sex on the measures of interest to this study.

Plasmids

pCAG-Cre was generated in the laboratory of F.P. as previously described (72). pAAV-CAG-FLEX-jGCaMP7b-WPRE was obtained from Addgene (#104497). pCAG-FLEX-eGFP was a gift from A. Maximov (Scripps). pAAV-CaMKIIa-bReaChes-mRuby3 was subcloned from pAAV-CaMKIIa-bReaChes-EYFP (gift from K. Deisseroth) by excising the EYFP sequence and replacing it with an mRuby3 fragment (cloned from pAAV-CAG-mRuby3-WPRE, Addgene #107744). pCAG-NES-jRGECO1a was generated by subcloning an NES-jRGECO1a insert (from pAAV-Syn-NES-jRGECO1a-WPRE, Addgene #100854) into a pCAG backbone. Subcloning was performed using the In-Fusion cloning kit (Takara Bio).

Single-cell electroporation

To electroporate single CA1PNs with custom plasmid combinations, animals were head-fixed under a two-photon microscope and anesthetized under a ketamine-xylazine mixture with continuous heat source. Sterile 1× phosphate-buffered saline was used to make a grounded solution in the bowl-shaped dental acrylic cap atop the imaging implant. A pipette containing fluorescent plasmid solution was lowered at an angle using a motorized micromanipulator (Scientifica) through the rectangular slit in the imaging coverslip and through the thin silicone protective layer while maintaining positive pipette pressure to avoid clogging. Once in the brain, the pipette was gradually lowered to the CA1 pyramidal layer (Fig. 1A, middle) while continuously expelling low amounts of fluorescent solution to aid visualization under 920-nm, two-photon excitation. As the pipette approached a putative CA1PN, tip resistance was used to monitor proximity. After a small deflection in resistance and while maintaining positive pressure, a train of negative voltage pulses was delivered from a stimulus isolator (AMPI) to the amplifier head stage using a custom-made switch circuit. Successful electroporation was confirmed based on a cell filling with Alexa Fluor dye and retaining the dye after careful pipette retraction (Fig. 1A, middle). Plasmid DNA expression was checked 48 hours later. To ensure successful genetic KO of *Pdzd8* (fig. S1, A to F) and clearance of PDZD8 protein (fig. S1, G and H), imaging experiments were conducted no less than 7 days after electroporation for either $Pdzd8^{F/F}$ or WT control mice. See the supplementary materials and methods for additional details.

Animal behavior

Mice ran along a spatially cued, 2-m treadmill belt under "random foraging" conditions in which water reward was delivered at a pseudorandom location once per lap. Initial delivery was nonoperant and was followed by a 3-s period during which the animal could receive an additional operant water reward by licking with a 50% reward rate. Mice were habituated on the same cued belts used for experiments for 1 to 2 days with training sessions under the in vivo imaging setup before data acquisition. For each mouse, the belt was never changed across days within a given experiment and was calibrated to the exact same length each day using a fixed landmark of blackout tape, an infrared beam, and custom software.

In vivo two-photon Ca^{2+} imaging

All imaging experiments described in the figures were performed using a two-photon 8-kHz resonant scanning microscope (Bruker) equipped with a Chameleon Ultra II Ti:sapphire laser (Coherent), tuned to 920 nm for green wavelength excitation, and a Fidelity-2 (Coherent) laser fixed at 1070 nm for red excitation. Excitation pathways were separately controlled and combined at the microscope. Green and red fluorescence were separated with an emission filter set (HQ525/70m-2p, HQ607/45m-2p, 575dcxr, Chroma Technology) and collected using channel-dedicated GaAsP (7422P-40, Hamamatsu Photonics) photomultiplier tubes (PMTs). A custom dual-stage preamplifier (1.4 × 105 dB, Bruker) was used to amplify signals before digitization. For details on the spine-level imaging experiments described in fig. S6, G to J, and movie S1, see the supplementary materials and methods.

Single-cell optogenetic PF induction

Induction sessions constituted the first imaging day of each single-cell imaging experiment. After a baseline period of five laps, a 1-s LED photostimulation was triggered at a fixed location along the treadmill belt for five consecutive laps. If preexisting somatic spatial tuning was evident (as visually assessed offline before baseline periods), LED location was set to be as far away as possible from the existing putative PF. To deliver LED photostimulation while simultaneously acquiring jGCaMP7b Ca^{2+} dynamics, 620-nm light from a C-mounted ultra-fast, ultra-high-powered LED (UHP-T-620-SR, Prizmatix) was passed through a dichroic mirror, allowing red light to pass into the objective lens back aperture while deflecting emitted green photons to their dedicated PMT (Fig. 1B, left). The red channel used to acquire mRuby3 signals was manually switched off during photostimulation. LED was triggered by a custom "Pockels-blanking" circuit that relayed an inverted Pockels cell blanking signal that was briefly activated during Y-galvonometer flyback periods and during toggling of the piezoelectric device. This approach allowed for high-powered (30 to 40 mW at sample), pulsed LED stimulation during image acquisition while protecting PMTs from potentially saturating LED photons that might otherwise be incompletely deflected from the green collection channel given high LED power. After optogenetic PF induction, the remainder of the 20-min imaging session was used to acquire postinduction data. All cells were tracked for 72 hours, except for two WT cells that died at the 72-hour mark (Fig. 5E and table S1). Of the 16 induction experiments performed, all but three sessions involved one CA1PN. Two WT experiments and one KO experiment involved two electroporated CA1PNs in the same field of view.

Detecting spontaneous PF formation events

To detect de novo PF formation events, we quantified the number of events falling within a search window ("hits") that corresponded to the PF bounds plus a forward extension of 20 cm to account for an expected backward shift in PF activity relative to formation location ([33]). We then used an $N = 2$ Gaussian mixture model (GMM) to initially differentiate between putative "PRE" and "POST" laps relative to PF formation. For each detected PF, the GMM was fit to a two-feature matrix with columns representing (i) a rolling average (weighted by Bartlett window) of hit counts by lap and (ii) lap numbers multiplied by exponentially decaying weights. The GMM fit was randomly initiated 1000 times for each PF, and the best fit was taken. The PF was assumed to have existed before session start if (i) the model did not converge, (ii) all laps were assigned to a single component, or (iii) the mean hit/lap rate in PRE laps was >1. These events were discarded. Next, we searched for a large Ca^{2+} transient (≥1 SD above mean amplitude) representing a putative PF-forming plateau potential. If a candidate transient was identified within 10 laps on either side of the first POST lap, then the first such event was taken as the PF-forming event. If no such candidate event was identified, then the putative de novo PF was discarded. To validate detected PF formation events, five conditions were imposed: (i) the same cell did not express an overlapping PF on the previous day, (ii) the hit rates in PRE versus POST laps were significantly different by the Mann-Whitney *U* test, (iii) the place cell fired within its PF on at least a third of POST laps, (iv) place cell events showed at least 35% specificity to the PF after formation, and (v) the PF did not form earlier than lap 10 to further mitigate the possibility that it had formed before imaging. For each PF-forming Ca^{2+} transient, we documented the corresponding lap, 2-cm spatial bin, and the imaging frame at which the event reached peak amplitude for future analyses. See fig. S14A for example performance and the supplementary materials and methods for additional details and rationale.

Analyzing PF formation events

For optogenetically induced PFs in single cells, an "LED zone" was defined to operationalize what activity should be considered induced. The LED zone was defined as a symmetric window, with a 15 cm radius, centered on LED onset position (Fig. 5C, green shaded area) on a cued, 2-m treadmill belt. For each induction experiment, spatial tuning heatmaps were calculated (as described in the supplementary materials and methods) based on preinduction (baseline), peri-induction, and postinduction laps using dF/F_0. Induction efficacy was calculated as the maximum difference between postinduction and baseline TCs within the LED zone. Baseline and postinduction TCs to the maximum observed LED response for a given ROI are shown in Fig. 5C. An ROI was considered "induced" if its mean postinduction dF/F_0 exceeded baseline dF/F_0 in at least one position bin (50 4-cm bins were used) within the LED zone. Periformation velocity (figs. S4C and S11, B and D) was defined as the mean velocity observed from 1.5 s before and 1.0 s after LED onset during the first induction lap. This temporal window was formulated to capture the bulk of the integral of the asymmetric plasticity kernel underlying BTSP (*32*). For spontaneously occurring PF formation events in the population imaging data, we adjusted our calculation of periformation velocity to accommodate the inherent uncertainty in estimating plateau potential onset times based on somatic GCaMP signals. Periformation velocity was calculated as the peak velocity within the spatial search window described above on the formation lap. See the supplementary materials and methods for additional details and rationale.

To calculate PF peak shifts, we increased the resolution of spatial bins to 0.5 cm (400 spatial bins per 2-m lap). Large, PF-forming Ca^{2+} transients were localized to the nearest 0.5-cm spatial bin based on the imaging frame corresponding to peak transient amplitude. Postformation TCs were calculated from all laps subsequent to the formation lap within an imaging session. Peak shifts were defined as the distance between peaks of formation and postformation TCs. Widths of de novo PFs were calculated using events from all laps.

REFERENCES AND NOTES

1. P. L. Greer, M. E. Greenberg, From synapse to nucleus: Calcium-dependent gene transcription in the control of synapse development and function. *Neuron* **59**, 846–860 (2008). doi: 10.1016/j.neuron.2008.09.002; pmid: 18817726
2. G. J. Stuart, N. Spruston, Dendritic integration: 60 years of progress. *Nat. Neurosci.* **18**, 1713–1721 (2015). doi: 10.1038/nn.4157; pmid: 26605882
3. M. Häusser, N. Spruston, G. J. Stuart, Diversity and dynamics of dendritic signaling. *Science* **290**, 739–744 (2000). doi: 10.1126/science.290.5492.739; pmid: 11052929
4. E. Nanou, W. A. Catterall, Calcium channels, synaptic plasticity, and neuropsychiatric disease. *Neuron* **98**, 466–481 (2018). doi: 10.1016/j.neuron.2018.03.017; pmid: 29723500
5. H. Takechi, J. Eilers, A. Konnerth, A new class of synaptic response involving calcium release in dendritic spines. *Nature* **396**, 757–760 (1998). doi: 10.1038/25547; pmid: 9874373
6. E. G. Gray, Axo-somatic and axo-dendritic synapses of the cerebral cortex: An electron microscope study. *J. Anat.* **93**, 420–433 (1959). pmid: 13829103
7. A. Peters, S. L. Palay, H. deF. Webster, *Fine Structure of the Nervous System: Neurons and Their Supporting Cells* (Oxford Univ. Press, 1991).
8. J. Spacek, K. M. Harris, Three-dimensional organization of smooth endoplasmic reticulum in hippocampal CA1 dendrites and dendritic spines of the immature and mature rat. *J. Neurosci.* **17**, 190–203 (1997). doi: 10.1523/JNEUROSCI.17-01-00190.1997; pmid: 8987748
9. Y. Hirabayashi *et al.*, ER-mitochondria tethering by PDZD8 regulates Ca^{2+} dynamics in mammalian neurons. *Science* **358**, 623–630 (2017). doi: 10.1126/science.aan6009; pmid: 29097544
10. T. Nakamura, J.-G. Barbara, K. Nakamura, W. N. Ross, Synergistic release of Ca2+ from IP3-sensitive stores evoked by synaptic activation of mGluRs paired with backpropagating action potentials. *Neuron* **24**, 727–737 (1999). doi: 10.1016/S0896-6273(00)81125-3; pmid: 10595522
11. T. Nakamura *et al.*, Inositol 1,4,5-trisphosphate (IP_3)-mediated Ca^{2+} release evoked by metabotropic agonists and backpropagating action potentials in hippocampal CA1 pyramidal neurons. *J. Neurosci.* **20**, 8365–8376 (2000). doi: 10.1523/JNEUROSCI.20-22-08365.2000; pmid: 11069943
12. T. Nakamura, N. Lasser-Ross, K. Nakamura, W. N. Ross, Spatial segregation and interaction of calcium signalling mechanisms in rat hippocampal CA1 pyramidal neurons. *J. Physiol.* **543**, 465–480 (2002). doi: 10.1113/jphysiol.2002.020362; pmid: 12205182
13. M. E. Larkum, S. Watanabe, T. Nakamura, N. Lasser-Ross, W. N. Ross, Synaptically activated Ca^{2+} waves in layer 2/3 and layer 5 rat neocortical pyramidal neurons. *J. Physiol.* **549**, 471–488 (2003). doi: 10.1113/jphysiol.2002.037614; pmid: 12692172
14. W. N. Ross, T. Nakamura, S. Watanabe, M. Larkum, N. Lasser-Ross, Synaptically activated ca2+ release from internal stores in CNS neurons. *Cell. Mol. Neurobiol.* **25**, 283–295 (2005). doi: 10.1007/s10571-005-3060-0; pmid: 16047542
15. M. Nishiyama, K. Hong, K. Mikoshiba, M.-M. Poo, K. Kato, Calcium stores regulate the polarity and input specificity of synaptic modification. *Nature* **408**, 584–588 (2000). doi: 10.1038/35046067; pmid: 11117745
16. K. F. Lee, C. Soares, J. P. Thivierge, J. C. Béïque, Correlated synaptic inputs drive dendritic calcium amplification and cooperative plasticity during clustered synapse development. *Neuron* **89**, 784–799 (2016). doi: 10.1016/j.neuron.2016.01.012; pmid: 26853305
17. F. W. Johenning *et al.*, Ryanodine receptor activation induces long-term plasticity of spine calcium dynamics. *PLOS Biol.* **13**, e1002181 (2015). doi: 10.1371/journal.pbio.1002181; pmid: 26098891
18. S. Bardo, M. G. Cavazzini, N. Emptage, The role of the endoplasmic reticulum Ca2+ store in the plasticity of central neurons. *Trends Pharmacol. Sci.* **27**, 78–84 (2006). doi: 10.1016/j.tips.2005.12.008; pmid: 16412523
19. N. Emptage, T. V. Bliss, A. Fine, Single synaptic events evoke NMDA receptor-mediated release of calcium from internal stores in hippocampal dendritic spines. *Neuron* **22**, 115–124 (1999). doi: 10.1016/S0896-6273(00)80683-2; pmid: 10027294
20. K. Svoboda, Z. F. Mainen, Synaptic [Ca2+]: Intracellular stores spill their guts. *Neuron* **22**, 427–430 (1999). doi: 10.1016/S0896-6273(00)80698-4; pmid: 10197523
21. C. R. Rose, A. Konnerth, Stores not just for storage. intracellular calcium release and synaptic plasticity. *Neuron* **31**, 519–522 (2001). doi: 10.1016/S0896-6273(01)00402-0; pmid: 11545711
22. A. Handler *et al.*, Distinct dopamine receptor pathways underlie the temporal sensitivity of associative learning. *Cell* **178**, 60–75.e19 (2019). doi: 10.1016/j.cell.2019.05.040; pmid: 31230716
23. A. H. Sharp *et al.*, Differential immunohistochemical localization of inositol 1,4,5-trisphosphate- and ryanodine-sensitive Ca2+ release channels in rat brain. *J. Neurosci.* **13**, 3051–3063 (1993). doi: 10.1523/JNEUROSCI.13-07-03051.1993; pmid: 8392539
24. M. Migliore, M. Ferrante, G. A. Ascoli, Signal propagation in oblique dendrites of CA1 pyramidal cells. *J. Neurophysiol.* **94**, 4145–4155 (2005). doi: 10.1152/jn.00521.2005; pmid: 16293591
25. M. Andreasen, J. D. C. Lambert, Factors determining the efficacy of distal excitatory synapses in rat hippocampal CA1 pyramidal neurones. *J. Physiol.* **507**, 441–462 (1998). doi: 10.1111/j.1469-7793.1998.441bt.x; pmid: 9518704
26. J. C. Magee, Dendritic hyperpolarization-activated currents modify the integrative properties of hippocampal CA1 pyramidal neurons. *J. Neurosci.* **18**, 7613–7624 (1998). doi: 10.1523/JNEUROSCI.18-19-07613.1998; pmid: 9742133
27. G. K. Pyapali, A. Sik, M. Penttonen, G. Buzsaki, D. A. Turner, Dendritic properties of hippocampal CA1 pyramidal neurons in the rat: Intracellular staining in vivo and in vitro. *J. Comp. Neurol.* **391**, 335–352 (1998). doi: 10.1002/(SICI)1096-9861(19980216)391:3<335::AID-CNE4>3.0.CO;2-2; pmid: 9492204
28. M. T. Harnett, J. K. Makara, N. Spruston, W. L. Kath, J. C. Magee, Synaptic amplification by dendritic spines enhances input cooperativity. *Nature* **491**, 599–602 (2012). doi: 10.1038/nature11554; pmid: 23103868
29. N. Spruston, Pyramidal neurons: Dendritic structure and synaptic integration. *Nat. Rev. Neurosci.* **9**, 206–221 (2008). doi: 10.1038/nrn2286; pmid: 18270515
30. 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).
31. J. O'Keefe, L. Nadel, *The Hippocampus As a Cognitive Map* (Clarendon Press, 1978).
32. K. C. Bittner *et al.*, Conjunctive input processing drives feature selectivity in hippocampal CA1 neurons. *Nat. Neurosci.* **18**, 1133–1142 (2015). doi: 10.1038/nn.4062; pmid: 26167906
33. K. C. Bittner, A. D. Milstein, C. Grienberger, S. Romani, J. C. Magee, Behavioral time scale synaptic plasticity underlies CA1 place fields. *Science* **357**, 1033–1036 (2017). doi: 10.1126/science.aan3846; pmid: 28883072
34. J. C. Magee, C. Grienberger, Synaptic plasticity forms and functions. *Annu. Rev. Neurosci.* **43**, 95–117 (2020). doi: 10.1146/annurev-neuro-090919-022842; pmid: 32075520
35. J. de Juan-Sanz *et al.*, Axonal endoplasmic reticulum Ca^{2+} content controls release probability in CNS nerve terminals. *Neuron* **93**, 867–881.e6 (2017). doi: 10.1016/j.neuron.2017.01.010; pmid: 28162809
36. N. B. Danielson *et al.*, Sublayer-specific coding dynamics during spatial navigation and learning in hippocampal area CA1. *Neuron* **91**, 652–665 (2016). doi: 10.1016/j.neuron.2016.06.020; pmid: 27397517
37. J. D. Zaremba *et al.*, Impaired hippocampal place cell dynamics in a mouse model of the 22q11.2 deletion. *Nat. Neurosci.* **20**, 1612–1623 (2017). doi: 10.1038/nn.4634; pmid: 28869582
38. Y. Ziv *et al.*, Long-term dynamics of CA1 hippocampal place codes. *Nat. Neurosci.* **16**, 264–266 (2013). doi: 10.1038/nn.3329; pmid: 23396101
39. J. B. Priestley, J. C. Bowler, S. V. Rolotti, S. Fusi, A. Losonczy, Signatures of rapid synaptic learning in the hippocampus during novel experiences [Preprint]. bioRxiv 450956 (2021); doi: 10.1101/2021.07.02.450956
40. C. Grienberger, J. C. Magee, Entorhinal cortex directs learning-related changes in CA1 representations [Preprint]. bioRxiv 472158 (2021); doi: 10.1101/2021.12.10.472158
41. M. E. Sheffield, D. A. Dombeck, Calcium transient prevalence across the dendritic arbour predicts place field properties. *Nature* **517**, 200–204 (2015). doi: 10.1038/nature13871; pmid: 25363782
42. M. E. J. Sheffield, M. D. Adoff, D. A. Dombeck, Increased prevalence of calcium transients across the dendritic arbor during place field formation. *Neuron* **96**, 490–504.e5 (2017). doi: 10.1016/j.neuron.2017.09.029; pmid: 29024668
43. S. V. Rolotti, H. Blockus, F. T. Sparks, J. B. Priestley, A. Losonczy, Reorganization of CA1 dendritic dynamics by hippocampal sharp-wave ripples during learning. *Neuron* S0896-6273(21)01034-5 (2022). doi: 10.1016/j.neuron.2021.12.017; pmid: 35041805
44. P. J. Dittmer, A. R. Wild, M. L. Dell'Acqua, W. A. Sather, STIM1 Ca^{2+} sensor control of L-type Ca^{2+}-channel-dependent dendritic spine structural plasticity and nuclear signaling. *Cell Rep.* **19**, 321–334 (2017). doi: 10.1016/j.celrep.2017.03.056; pmid: 28402855
45. N. Holbro, A. Grunditz, T. G. Oertner, Differential distribution of endoplasmic reticulum controls metabotropic signaling and plasticity at hippocampal synapses. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 15055–15060 (2009). doi: 10.1073/pnas.0905110106; pmid: 19706463
46. M. Dur-e-Ahmad, M. Imran, A. Gul, Calcium dynamics in dendritic spines: A link to structural plasticity. *Math. Biosci.* **230**, 55–66 (2011). doi: 10.1016/j.mbs.2011.01.002; pmid: 21295598
47. P. J. Dittmer, M. L. Dell'Acqua, W. A. Sather, Synaptic crosstalk conferred by a zone of differentially regulated Ca^{2+} signaling in the dendritic shaft adjoining a potentiated spine. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 13611–13620 (2019). doi: 10.1073/pnas.1902461116; pmid: 31209051
48. E. A. Finch, G. J. Augustine, Local calcium signalling by inositol-1,4,5-trisphosphate in Purkinje cell dendrites. *Nature* **396**, 753–756 (1998). doi: 10.1038/25541; pmid: 9874372
49. S. Druckmann *et al.*, Structured synaptic connectivity between hippocampal regions. *Neuron* **81**, 629–640 (2014). doi: 10.1016/j.neuron.2013.11.026; pmid: 24412418
50. T. Kleindienst, J. Winnubst, C. Roth-Alpermann, T. Bonhoeffer, C. Lohmann, Activity-dependent clustering of functional

synaptic inputs on developing hippocampal dendrites. *Neuron* **72**, 1012–1024 (2011). doi: 10.1016/j.neuron.2011.10.015; pmid: 22196336

51. O. Kwon, L. Feng, S. Druckmann, J. Kim, Schaffer collateral inputs to CA1 excitatory and inhibitory neurons follow different connectivity rules. *J. Neurosci.* **38**, 5140–5152 (2018). doi: 10.1523/JNEUROSCI.0155-18.2018; pmid: 29728449
52. N. Takahashi *et al*., Locally synchronized synaptic inputs. *Science* **335**, 353–356 (2012). doi: 10.1126/science.1210362; pmid: 22267814
53. M. D. Adoff *et al*., The functional organization of excitatory synaptic input to place cells. *Nat. Commun.* **12**, 3558 (2021). doi: 10.1038/s41467-021-23829-y; pmid: 34117238
54. H. Dana *et al*., High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. *Nat. Methods* **16**, 649–657 (2019). doi: 10.1038/s41592-019-0435-6; pmid: 31209382
55. C. Dong, A. D. Madar, M. E. J. Sheffield, Distinct place cell dynamics in CA1 and CA3 encode experience in new environments. *Nat. Commun.* **12**, 2977 (2021). doi: 10.1038/s41467-021-23260-3; pmid: 34016996
56. T. Hainmueller, M. Bartos, Parallel emergence of stable and dynamic memory engrams in the hippocampus. *Nature* **558**, 292–296 (2018). doi: 10.1038/s41586-018-0191-2; pmid: 29875406
57. N. R. Kinsky, D. W. Sullivan, W. Mau, M. E. Hasselmo, H. B. Eichenbaum, Hippocampal Place Fields Maintain a Coherent and Flexible Map across Long Timescales. *Curr. Biol.* **28**, 3578–3588.e6 (2018). doi: 10.1016/j.cub.2018.09.037; pmid: 30393037
58. N. L. Golding, W. L. Kath, N. Spruston, Dichotomy of action-potential backpropagation in CA1 pyramidal neuron dendrites. *J. Neurophysiol.* **86**, 2998–3010 (2001). doi: 10.1152/jn.2001.86.6.2998; pmid: 11731556
59. M. Kano, O. Garaschuk, A. Verkhratsky, A. Konnerth, Ryanodine receptor-mediated intracellular calcium release in rat cerebellar Purkinje neurones. *J. Physiol.* **487**, 1–16 (1995). doi: 10.1113/jphysiol.1995.sp020857; pmid: 7473240
60. H. Takahashi, J. C. Magee, Pathway interactions and synaptic plasticity in the dendritic tuft regions of CA1 pyramidal neurons. *Neuron* **62**, 102–111 (2009). doi: 10.1016/j.neuron.2009.03.007; pmid: 19376070
61. A. Losonczy, J. K. Makara, J. C. Magee, Compartmentalized dendritic plasticity and input feature storage in neurons. *Nature* **452**, 436–441 (2008). doi: 10.1038/nature06725; pmid: 18368112
62. J. S. Lee, J. J. Briguglio, J. D. Cohen, S. Romani, A. K. Lee, The statistical structure of the hippocampal code for space as a function of time, context, and value. *Cell* **183**, 620–635.e22 (2020). doi: 10.1016/j.cell.2020.09.024; pmid: 33035454
63. E. Campanac, D. Debanne, Spike timing-dependent plasticity: A learning rule for dendritic integration in rat CA1 pyramidal neurons. *J. Physiol.* **586**, 779–793 (2008). doi: 10.1113/jphysiol.2007.147017; pmid: 18048448
64. K. Diba, G. Buzsáki, Hippocampal network dynamics constrain the time lag between pyramidal cells across modified environments. *J. Neurosci.* **28**, 13448–13456 (2008). doi: 10.1523/JNEUROSCI.3824-08.2008; pmid: 19074018
65. K. Mizuseki, S. Royer, K. Diba, G. Buzsáki, Activity dynamics and behavioral correlates of CA3 and CA1 hippocampal pyramidal neurons. *Hippocampus* **22**, 1659–1680 (2012). doi: 10.1002/hipo.22002; pmid: 22367959
66. O. Garaschuk, Y. Yaari, A. Konnerth, Release and sequestration of calcium by ryanodine-sensitive stores in rat hippocampal neurones. *J. Physiol.* **502**, 13–30 (1997). doi: 10.1111/j.1469-7793.1997.013bl.x; pmid: 9234194
67. S. A. Neymotin *et al*., Neuronal calcium wave propagation varies with changes in endoplasmic reticulum parameters: A computer model. *Neural Comput.* **27**, 898–924 (2015). doi: 10.1162/NECO_a_00712; pmid: 25734493
68. H. Jia, N. L. Rochefort, X. Chen, A. Konnerth, Dendritic organization of sensory input to cortical neurons in vivo. *Nature* **464**, 1307–1312 (2010). doi: 10.1038/nature08947; pmid: 20428163
69. X. Chen, U. Leischner, N. L. Rochefort, I. Nelken, A. Konnerth, Functional mapping of single spines in cortical neurons in vivo. *Nature* **475**, 501–505 (2011). doi: 10.1038/nature10193; pmid: 21706031
70. Z. Varga, H. Jia, B. Sakmann, A. Konnerth, Dendritic coding of multiple sensory inputs in single cortical neurons in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 15420–15425 (2011). doi: 10.1073/pnas.1112355108; pmid: 21876170
71. M. E. Sheffield, D. A. Dombeck, The binding solution? *Nat. Neurosci.* **18**, 1060–1062 (2015). doi: 10.1038/nn.4075; pmid: 26216459
72. J. Courchet *et al*., Terminal axon branching is regulated by the LKB1-NUAK1 kinase pathway via presynaptic mitochondrial capture. *Cell* **153**, 1510–1525 (2013). doi: 10.1016/j.cell.2013.05.021; pmid: 23791179
73. Data and code for: J. K. O'Hare, K. C. Gonzalez, S. A. Herrlinger, Y. Hirabayashi, V. L. Hewitt, H. Blockus, M. Szoboszlay, S. V. Rolotti, T. C. Geiller, A. Negrean, V. Chelur, F. Polleux, A. Losonczy, Compartment-specific tuning of dendritic feature selectivity by intracellular Ca^{2+} release, Zenodo (2022); doi: 10.5281/zenodo.6051319

ACKNOWLEDGMENTS

We thank S. Fusi for productive discussion regarding intradendritic dynamics, T. Tabachnik and Zuckerman Institute Advanced Instrumentation for design of a custom-threaded optics adapter used to target virally labeled cells for electrophysiological recordings, A. Villegas for assistance piloting an immunoassay to validate the conditional knockout strategy, and A. Nelson and M. Rossi for invaluable comments on the manuscript. **Funding:** This work was supported by the National Institutes of Health (grants R01MH100631, R01NS094668, and U19NS104590 to A.L.; grants R01NS067557 and R01NS094668 to F.P.; grant F32MH118716 to J.K.O.; grant K00NS105187 to S.A.H.; grant F31MH117892 to S.V.R.; grant K99NS115984 to H.B.; and grant T32NS064928 to K.C.G.); the Japan Science and Technology Agency (JST PRESTO grant JPMJPR16F7 to Y.H.); the Zegar Family Foundation (A.L.); and the Foundation Roger De Spoelberch (F.P.). **Author contributions:** Conceptualization: J.K.O., A.L., F.P.; Data curation: J.K.O., K.C.G., S.V.R., V.C.; Formal analysis: J.K.O., V.L.H.; Funding acquisition: J.K.O., A.L., F.P.; Investigation: J.K.O., K.C.G., S.A.H., V.L.H.; Methodology: J.K.O., K.C.G., S.A.H., S.V.R., H.B., M.S., T.C.G., A.N., V.L.H., V.C.; Project administration: A.L., F.P.; Resources: Y.H., S.A.H., H.B., M.S., K.C.G., A.L., F.P.; Software: J.K.O., S.V.R., V.C., T.C.G.; Supervision: A.L., F.P.; Validation: J.K.O., S.A.H., K.C.G., V.L.H., V.C.; Visualization: J.K.O., S.A.H., Y.H., V.L.H.; Writing – original draft: J.K.O., A.L., F.P.; Writing – review and editing: all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All plasmid DNA constructs generated in the course of this study are publicly available at Addgene. *Pdzd8*$^{F/F}$ mice are available upon request for academic, noncommercial use. Data and code are available at Zenodo (*73*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm1670
Materials and Methods
Figs. S1 to S15
Table S1
References (*74*–*89*)
Movie S1
MDAR Reproducibility Checklist

30 August 2021; accepted 14 February 2022
10.1126/science.abm1670

RESEARCH ARTICLE SUMMARY

EVOLUTION

Direct observation of adaptive tracking on ecological time scales in *Drosophila*

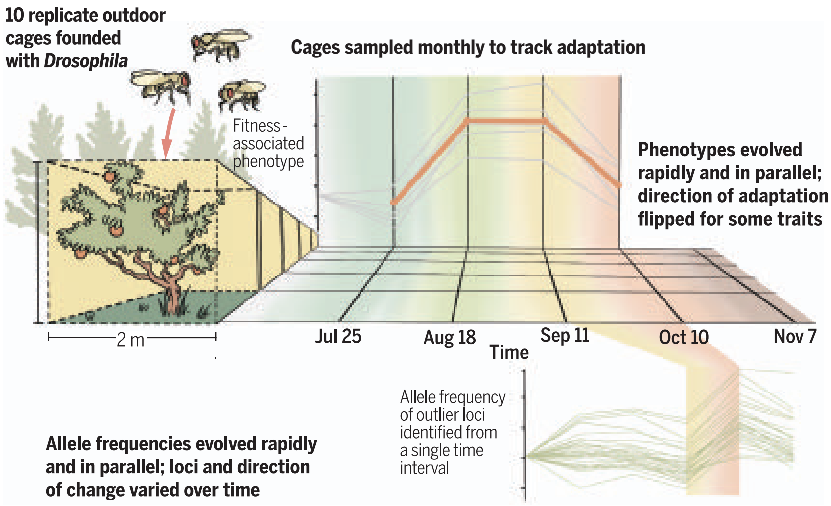
Seth M. Rudman*†, Sharon I. Greenblum*†, Subhash Rajpurohit†, Nicolas J. Betancourt, Jinjoo Hanna, Susanne Tilk, Tuya Yokoyama, Dmitri A. Petrov, Paul Schmidt*

INTRODUCTION: Adaptation is a foundational process in evolutionary biology that is central to human health and the conservation of biodiversity. Adaptive tracking, defined as continuous adaptation in response to rapid environmental change, is a potentially critical mechanism by which populations persist in changing environments. However, little is known about the pace, extent, and magnitude of adaptive tracking in response to natural temporal variation. This may be attributable to the historical perception that adaptation occurs slowly relative to environmental change as well as to difficulties in measurement. Replicated field experiments that directly measure phenotypic and genomic evolution in response to ongoing environmental change could provide the resolution necessary to answer fundamental questions about the process of adaptation, including those about adaptive tracking. If adaptive tracking does occur in populations experiencing rapid environmental change, it would alter our understanding of the importance of evolutionary change for ecological outcomes.

RATIONALE: The design of our field experiment focused on three key elements to study adaptive tracking in response to ongoing environmental change: (i) generating highly accurate measurements of phenotypic and genomic evolution, (ii) taking these measurements on a time scale similar to that of natural environmental change, and (iii) collecting measurements from independent replicate populations experiencing similar environmental conditions. Specifically, we directly measured phenotypic and genomic evolution in response to natural seasonal change across 10 independent replicate field populations, each comprising up to 100,000 *Drosophila melanogaster* individuals. We measured the evolution of heritable phenotypes related to stress tolerance and reproductive output under typical laboratory conditions (repeated common-garden rearing) over monthly intervals from July to November 2014. Individuals collected at the same intervals from each population were used for full-genome pooled sequencing to measure genomic change over time. We specifically assessed the degree of parallel phenotypic and genomic change across replicate populations, as any parallelism would be strong evidence of the deterministic evolutionary process of adaptation. We further looked for reversals in the direction of adaptation over time, which would demonstrate variation in the direction of natural selection (termed fluctuating selection) and would provide strong evidence for adaptive tracking.

RESULTS: We assayed phenotypic evolution of six fitness-associated phenotypes and genomic evolution at 1.9 million single-nucleotide polymorphic sites across the genome. We found clear evidence of parallel phenotypic and genomic evolution indicative of adaptive tracking. Phenotypic adaptation was pervasive and remarkably fast relative to measures of trait change from many prior studies. Moreover, the direction of evolution changed over time for development rate (a key component of reproductive output) and desiccation tolerance (an important trait in stress resistance). Similar patterns were observed in genomic data, with clear signatures of parallel allele frequency change across time intervals, including some that resulted from fluctuating selection. Genomic regions implicated in selection were found on all chromosomes, yet most were under selection over only a single time interval. At least 165 independent genomic regions were under selection during the experiment; allele frequencies at >60% of variant sites genome-wide were shaped in part by selection. Taken together, these data underscore that natural selection is capable of driving evolution in multiple fitness-associated phenotypes and much of the genome even over short time scales.

CONCLUSION: Our experiment shows the feasibility of observing adaptive tracking in response to environmental change in real time using replicated field experiments. Our data from this approach support a model of adaptive tracking in which populations adapt in response to continuous environmental change, with selection acting on multiple phenotypes and tens to hundreds of genetic variants. The pronounced action of fluctuating selection suggests that evolutionary rates may be systematically underestimated and that fluctuating selection could be an underappreciated mechanism that maintains diversity. Determining whether adaptive tracking is a general feature of natural populations, and elucidating the mechanisms by which it occurs, could be transformative for understanding the generation and maintenance of biodiversity. ■



Adaptation tracks ongoing environmental change. We conducted a replicated longitudinal field experiment using outbred *Drosophila* populations to assess the phenotypic and genomic basis of adaptation in response to natural seasonal environmental change. We observed fast phenotypic adaptation in multiple phenotypes, parallel change in allele frequencies at dozens of independent loci, and reversals in the direction of adaptative evolution for some traits and loci over monthly time intervals.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: seth.rudman@wsu.edu (S.M.R.); greensi@lbl.gov (S.I.G.); schmidtp@upenn.edu (P.S.)
†These authors contributed equally to this work.

Cite this article as S. M. Rudman *et al.*, *Science* **375**, eabj7484 (2022). DOI: 10.1126/science.abj7484

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abj7484

RESEARCH ARTICLE

EVOLUTION

Direct observation of adaptive tracking on ecological time scales in *Drosophila*

Seth M. Rudman[1,2]*†, Sharon I. Greenblum[3,4]*†, Subhash Rajpurohit[1,5]†, Nicolas J. Betancourt[1], Jinjoo Hanna[1], Susanne Tilk[3], Tuya Yokoyama[3], Dmitri A. Petrov[3], Paul Schmidt[1]*

Direct observation of evolution in response to natural environmental change can resolve fundamental questions about adaptation, including its pace, temporal dynamics, and underlying phenotypic and genomic architecture. We tracked the evolution of fitness-associated phenotypes and allele frequencies genome-wide in 10 replicate field populations of *Drosophila melanogaster* over 10 generations from summer to late fall. Adaptation was evident over each sampling interval (one to four generations), with exceptionally rapid phenotypic adaptation and large allele frequency shifts at many independent loci. The direction and basis of the adaptive response shifted repeatedly over time, consistent with the action of strong and rapidly fluctuating selection. Overall, we found clear phenotypic and genomic evidence of adaptive tracking occurring contemporaneously with environmental change, thus demonstrating the temporally dynamic nature of adaptation.

Continuous adaptation in response to rapidly changing environmental conditions, termed adaptive tracking, could be a crucial mechanism by which populations respond to environmental change. Adaptive tracking has historically received little study because of the impression that adaptive evolutionary change is too slow to track complex and rapidly changing selection pressures in the wild (*1*). Moreover, theory suggests that variable and complex selective pressures should in general lead to the evolution of phenotypic plasticity or bet-hedging (*2*, *3*). Yet multiple longitudinal field studies and experiments demonstrate that adaptation can indeed occur very rapidly at individual traits or loci in response to strong environmental perturbations (*4*–*10*). Whether this translates into populations undergoing adaptive tracking in response to multifarious ecological changes, when theory predicts that pleiotropy (cases where a single gene affects multiple traits) should constrain natural selection and prevent adaptive tracking (*11*, *12*), is unknown. If adaptive tracking does indeed occur in such situations, it would have broad implications for our understanding of the limits and pace of polygenic adaptation (*13*), the prevalence of fluctuating selection (*14*) and its role in the maintenance of genetic variation (*15*), and the importance of rapid adaptation in ecological outcomes (*16*).

Identification of adaptive tracking requires direct measurement of phenotypic and genotypic evolution across replicate field populations in response to ongoing natural environmental change. Ideally, an experimental system would provide (i) the means for highly accurate measurements of even subtle, heritable shifts in key independent fitness-related phenotypes and loci under selection; (ii) the ability to assay multiple replicate populations exhibiting some degree of ecological and environmental realism to detect parallel genetic and phenotypic changes indicative of adaptation (*17*); and (iii) high-resolution temporal sampling to quantify rapid fluctuations in the magnitude and direction of selection as environmental changes occur.

Here, we used field mesocosms to measure the extent, pace, repeatability, and genomic basis of adaptive tracking in *Drosophila melanogaster* within the naturally fluctuating, temperate environment of a single growing season in Pennsylvania, USA (*10*, *18*, *19*) (Fig. 1). The design precluded migration and allowed populations to expand to a large adult census size (on the order of 100,000 adults in each replicate at the maximum population size). To initiate the experiment, we derived an outbred baseline population of *D. melanogaster* from a set of 80 inbred strains originally collected in the spring from Pennsylvania (table S1). On 15 July 2014, 10 replicate cages were each founded with 1000 individuals from the baseline population. All populations were fed every second day and tracked until the first hard frost on 7 November 2014. Specifically, at four time points we measured the evolution of six complex, fitness-associated phenotypes, focusing on those related to either reproductive output (fecundity, egg size, and development rate) or stress tolerance (starvation tolerance, chill coma recovery, and desiccation tolerance) (Fig. 1). To do so, we repeatedly collected and reared individuals from each field cage in standard laboratory conditions (i.e., multigeneration common garden) to distinguish evolution from phenotypic plasticity. All phenotypes were measured in the F_3 generation. We also tracked changes in allele frequencies genome-wide in each replicate using pooled sequencing at five time points. We employed haplotype-based allele frequency estimation (*20*) to generate highly accurate allele frequency trajectories. We assessed changes in allele frequency at individual sites and in clusters of linked sites to assess the magnitude and genetic architecture of adaptation.

Phenotypic patterns of adaptation and adaptive tracking

Population dynamics were largely consistent among the replicates; population size increased sharply during summer, peaked in early fall, and then declined steadily as minimum daily temperatures declined in late fall (Fig. 1). These population dynamics mimic the patterns observed in *D. melanogaster* populations (*21*) and many other multivoltine organisms inhabiting temperate natural environments, with summer exponential growth, high densities in late summer to early fall, and late-fall population declines. Egg production showed a similar pattern (fig. S1), and overall recruitment from egg to adult was low (fig. S2). Similarity in the ecological conditions among replicate populations, including abiotic factors (fig. S3) and population dynamics (Fig. 1), suggests that similar selective landscapes may have driven parallel evolution across replicates.

Phenotypic evolution was rapid and parallel, but temporal patterns varied across traits. To measure phenotypic evolution, we sampled individuals from the founding population and ~2500 eggs from each cage at the first four time points (25 July, 18 August, 11 September, 10 October), reared them in common-garden laboratory conditions for three generations, and assayed phenotypes in the F_3 progeny (Fig. 1). For all six phenotypes, which are known to be polygenic and associated with fitness (*22*), we observed substantial trait evolution with an average of 23% change in the mean trait value for each cage across all phenotypes over each time interval. Variation in environmental parameters among cages did not implicate any individual factors as agents of selection (fig. S4); this may be attributable to the limited variation between cages or the complexity of the selective landscape. Prior experiments conducted in these mesocosms have found evidence of rapid adaptation in response to experimentally manipulated agents of divergent selection (*10*, *19*).

All six phenotypes showed evidence of parallel evolution, indicative of adaptation, over

---

[1]Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA. [2]School of Biological Sciences, Washington State University, Vancouver, WA 98686, USA. [3]Department of Biology, Stanford University, Stanford, CA 94305, USA. [4]Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [5]Department of Biological and Life Sciences, Ahmedabad University, Ahmedabad 380009, GJ, India.
***Corresponding author. Email: seth.rudman@wsu.edu (S.M.R.); greensi@lbl.gov (S.I.G.); schmidtp@upenn.edu (P.S.)**
†These authors contributed equally to this work.

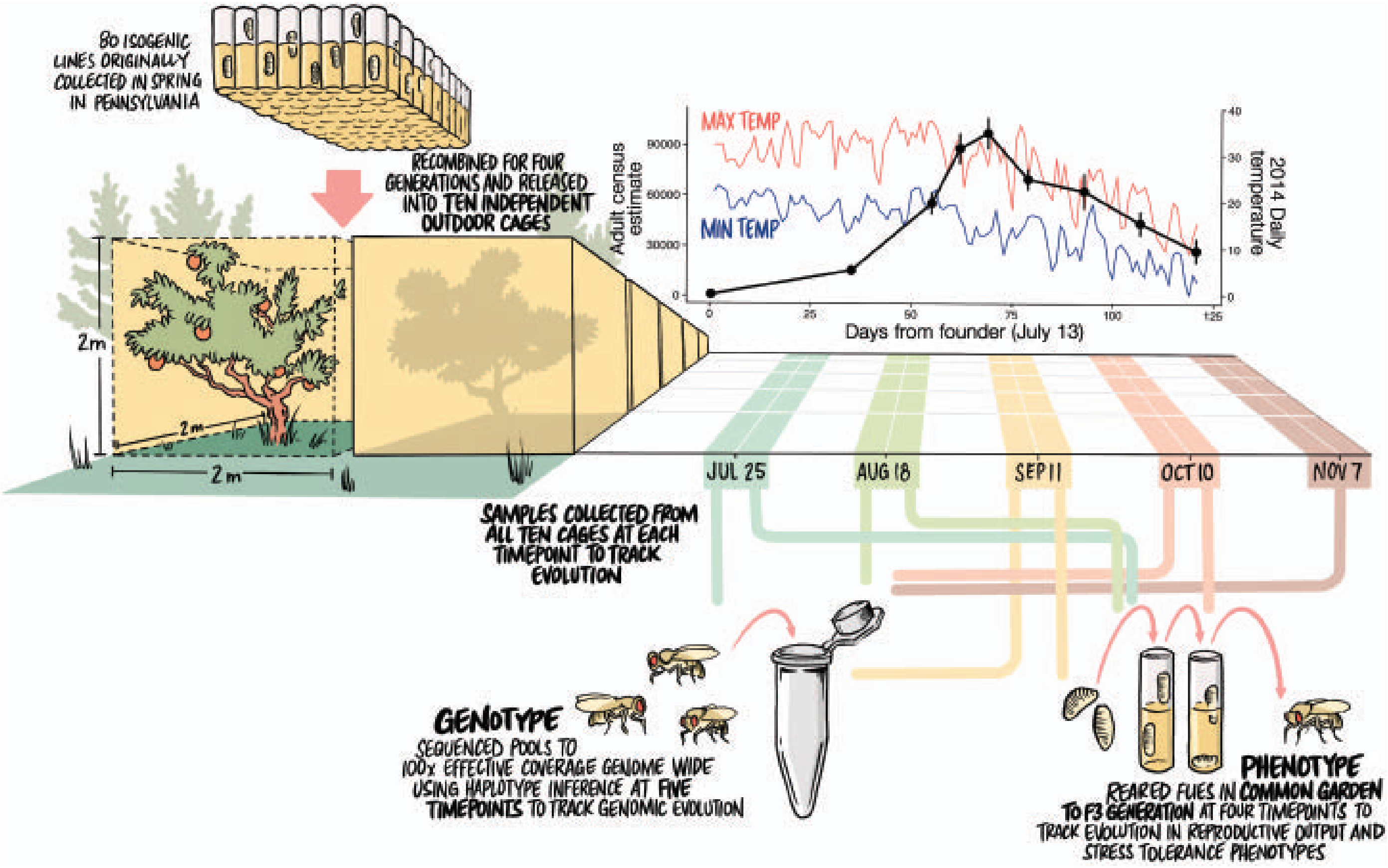


Fig. 1. Experimental arena, design, and population dynamics. The experiment was designed to reflect ecological and evolutionary realism while testing for adaptation using replicate populations. Eighty inbred lines originally collected in spring from an orchard in Pennsylvania were recombined and expanded for four generations into a genetically diverse outbred population in the laboratory. From this outbred population, we used 500 males and 500 females to found each of 10 independent outdoor cages (2 m × 2 m × 2 m). We measured daily minimum and maximum temperatures (blue and red lines, respectively) and estimated adult population size of each replicate over 4 months of seasonal change (black line, mean; gray lines, per replicate). To study adaptation, we tracked phenotypic evolution by collecting eggs from each replicate, rearing them in common-garden laboratory conditions for three generations, and then measuring six fitness-associated phenotypes. We conducted this procedure on the founder population and at four subsequent time points to measure phenotypic evolution over time. To study adaptation at the genomic level, we sequenced pools of 100 females from each cage to >100× effective coverage at five time points and assessed changes in allele frequencies.

time. Four of the six phenotypes evolved rapidly, repeatedly, and in a consistent direction across the duration of the experiment (fecundity, $F_{3,27} = 43.75$, $P < 0.0001$; egg size, $F_{3,27} = 11.5$, $P < 0.0001$; starvation, $F_{4,36} = 129.05$, $P < 0.0001$; chill coma recovery, $F_{4,36} = 197.75$, $P < 0.0001$) (Fig. 2). The magnitude of change was often substantial. For example, the average increase in fecundity was 61% over each monthly sampling interval across replicates, representing one to four overlapping generations. Desiccation tolerance and development rate also evolved rapidly and in parallel ($F_{4,36} = 86.66$, $P < 0.0001$, Fig. 2C; $F_{4,36} = 98.70$, $P < 0.0001$, Fig. 2F), but the direction of evolution varied over time. Fluctuations in the direction of evolution for these phenotypes had considerable effects on phenotypic trajectories; for desiccation tolerance, the amount of evolution measured over the whole experiment (founder to 10 October) was less than what was observed over the first interval (founder to 25 July). Identifying the fitness effects of any specific instance of phenotypic evolution is complicated by underlying correlations among traits, pleiotropy, and an unknown (and potentially temporally variable) phenotype-to-fitness map, but the pace and parallelism of phenotypic evolution are suggestive of strong links to fitness.

The pace of parallel trait evolution observed over the short time scales examined in this study was unusually fast. As expected, we observed rapid parallel evolution when outbred laboratory populations were introduced into the field enclosures and adapted to the field environment (founder → T_1). However, we also observed evidence of rapid adaptation between intervals in the enclosures for all six phenotypes, with some showing reversals in the direction of evolution across intervals (Fig. 2, C and F). The rate of phenotypic adaptation, calculated in haldanes [phenotypic evolution in units of standard deviations of the trait per generation (*23*, *24*)], was computed as a mean change across replicates for each phenotype over each interval and across the whole experiment (Fig. 2G). The rate of adaptation over the whole experiment ranged from moderate to extremely fast for different traits (0 to 0.8 haldanes) (*25*). However, when calculated over each sampling interval, the rate of adaptation was often comparable to or faster than the pace of phenotypic change measured in any prior field study or experiment (Fig. 2G).

The pace, magnitude, and parallelism of the phenotypic evolution we observed is notable for three reasons: (i) The evolutionary rates were calculated on the basis of the phenotypic shifts of the F_3 progeny in common-garden conditions, thus excluding phenotypic plasticity as the driver of change. (ii) Because we focused only on the parallel phenotypic shifts across the cages, our estimates describe the rate of putatively adaptive phenotypic change.

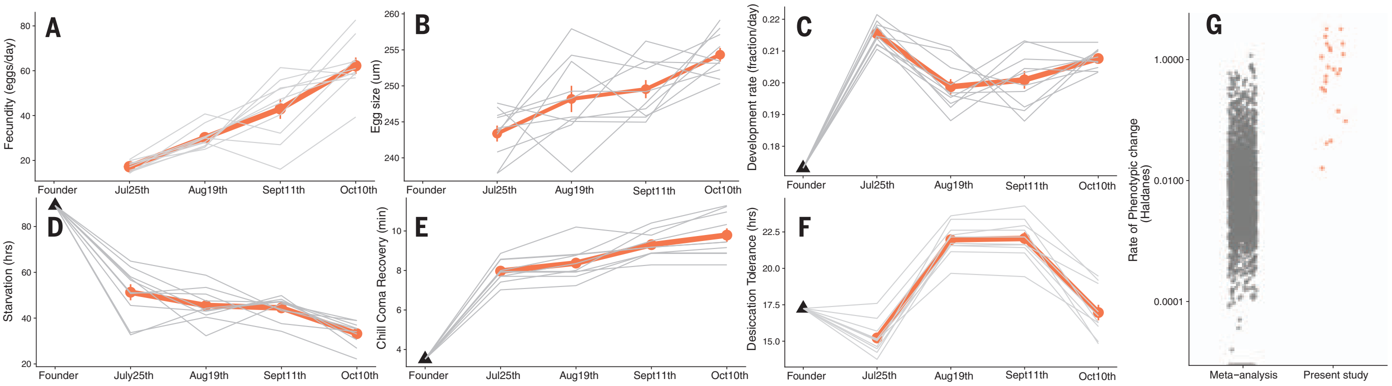


Fig. 2. Parallel evolution of stress tolerance traits, reproductive output traits, and comparison of the rate of adaptation. (**A** to **F**) Trajectories of phenotypic evolution for reproductive-associated traits [(A) to (C)] and stress resistance traits [(D) to (F)] as measured after three generations of common-garden rearing. (A) Mean fecundity as number of eggs per female per day. (B) Mean egg size. (C) Development rate as the fraction of development to pupation completed in 1 day [1/(total hours/24)]. (D) Starvation tolerance as time to death by starvation. (E) Recovery time after chill coma. (F) Desiccation tolerance as time to death from desiccation. Black points are the mean phenotypes of the founding population, gray lines represent mean phenotypic trajectories of individual populations, and red lines are the mean of all cage means. (**G**) Comparison of the rates of adaptation from this experiment over individual intervals (red) to rates of phenotypic change from a prior meta-analysis (gray) (*25*).

(iii) These patterns of rapid adaptation were observed for multiple fitness-associated phenotypes, each with a complex and likely distinct genetic architecture (*26*). Overall, our results show that strong and temporally variable natural selection can consistently drive rapid and polygenic adaptation of multiple fitness-associated phenotypes on the same time scale as the environmental change.

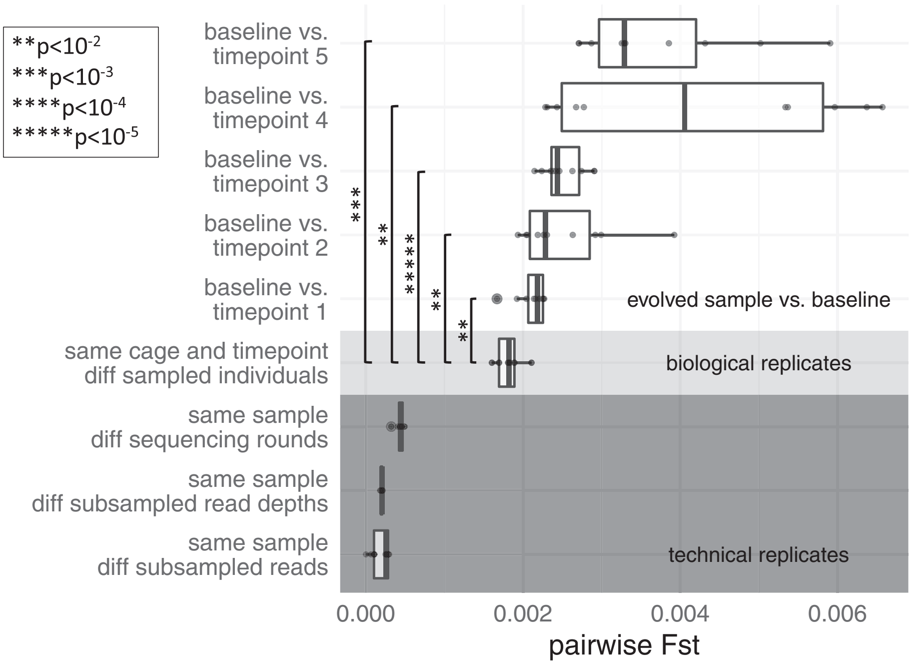


Fig. 3. Genome-wide evidence of evolutionary divergence. Distributions of genome-wide mean pairwise F_{ST} values between technical replicates (dark gray; same flies, different reads), biological replicates (light gray; different flies, same time point), and experimental samples from different time points compared to baseline (white). Note that negligible F_{ST} values between pairs of technical replicates are consistent with extreme precision of haplotype-derived allele frequencies (HAFs), which suggests that the variance in allele frequency estimates for biological replicates is primarily driven by sampling of different individuals. Asterisks represent the significance of divergence over time relative to biological replicates (t test).

Genomic patterns of rapid adaptation and adaptive tracking

To investigate the genomic architecture underlying the observed rapid phenotypic adaptation, we performed whole-genome pooled sequencing of 100 randomly selected individuals from the baseline population and from each replicate population at five time points across the experiment (Fig. 1). Allele frequencies at 1.9 million biallelic sites were inferred for each sample via haplotype inference using HAF-pipe (*20*) at accuracy levels consistent with an "effective coverage" of >100× (fig. S5 and table S2) (*27*). This high-resolution dataset yielded strong evidence for rapid genome-wide evolution. Specifically, we observed that the genome-wide estimates of F_{ST} between the founder population and all five monthly time points (mean 0.0030 ± 0.0002 standard error) exceeded expected margins of error based on technical and biological replicates (0.00026 ± 0.000024 and 0.0018 ± 0.000048, respectively, $P < 2 \times 10^{-8}$, t test; Fig. 3). Furthermore, divergence from the founder population changed significantly over time, both genome-wide ($P < 2.3 \times 10^{-5}$, Kruskal-Wallace P value for difference in means across time points) and for individual chromosomes ($P < 0.006$; fig. S6). Given the large population sizes (up to 10^5), it is unlikely that such substantial evolutionary change can be attributed solely to random genetic drift.

Further examination of the magnitude and direction of evolution across the 10 replicate cages showed substantial genomic adaptation, as defined by parallel, and thus deterministic, allele frequency shifts across replicate cages. To test for parallel shifts, we used a leave-one-out cross-validation approach. For each monthly time interval ($T_i \rightarrow T_{i+1}$, $i = 1, 2, 3, 4$), we used a generalized linear model (GLM) to identify sets of single-nucleotide polymorphisms (SNPs) whose frequency shifted significantly across the nine training cages, and then tested whether shifts at those SNPs in the 10th left-out cage exceeded shifts at randomly chosen matched control sites. Using this test, we found widespread parallel genomic adaptation for the first three sampling intervals (in 28 of 30 leave-one-out tests) (Fig. 4B). The pattern of parallelism was muted and evolution was more idiosyncratic in $T_4 \rightarrow T_5$. We also repeated the procedure for SNPs that shifted across the whole experiment ($T_1 \rightarrow T_5$) and found a similarly strong signal of parallel adaptation (10 of 10 tests). The magnitude of allele frequency shifts in each interval (2 to 8%) and over the whole experiment (2 to 5%) corresponded to very strong effective selection strength at the most parallel sites of ~10 to 50% per monthly interval (one to four generations) (27). This pattern was largely repeated when analyzing sites from each chromosome individually (fig. S7). In simulated populations with the same demographics as the experimental populations, allele frequency shifts of this magnitude were consistently achieved with selection coefficients of ≤50% on alleles spanning a wide range of initial frequencies over similar time scales (table S3) (27). The pronounced parallel shifts in allele frequency across independent populations demonstrate the strong action of natural selection.

Our cross-validation analysis also yielded clear evidence of variation in the magnitude and direction of selection over time, consistent with the observed patterns of phenotypic evolution for some traits (Fig. 2). Specifically, the leave-one-out analysis and the time-series genomic data allowed us to examine the full trajectory of alleles detected at any specific time interval (T_{det}). We found that these alleles often shifted significantly more than alleles at control sites (Fig. 4B) at other time intervals; however, the nature of these shifts varied over time. In some left-out cages and at some time intervals, alleles shifted in a direction consistent with their behavior during T_{det} (orange points); however, in other cases the direction flipped, resulting in significant reverse shifts (blue points). Reverse shifts were strongest for sites with $T_{\mathrm{det}} = T_3 \rightarrow T_4$ (August → September) during the time when populations expanded most rapidly and reached their maximum. These $T_3 \rightarrow T_4$ parallel sites showed consistent shifts in the opposite direction during the preceding interval ($T_2 \rightarrow T_3$, July → August) when the populations were still expanding. In many cages, these sites also shifted in the opposite direction during the subsequent ($T_4 \rightarrow T_5$, October → November) interval when population sizes were declining. These patterns likely reflect the action of rapidly fluctuating selection over the 4 months of the experiment.

With a complex and rapidly fluctuating selective landscape, adaptation occurs over multiple time scales simultaneously, with clear

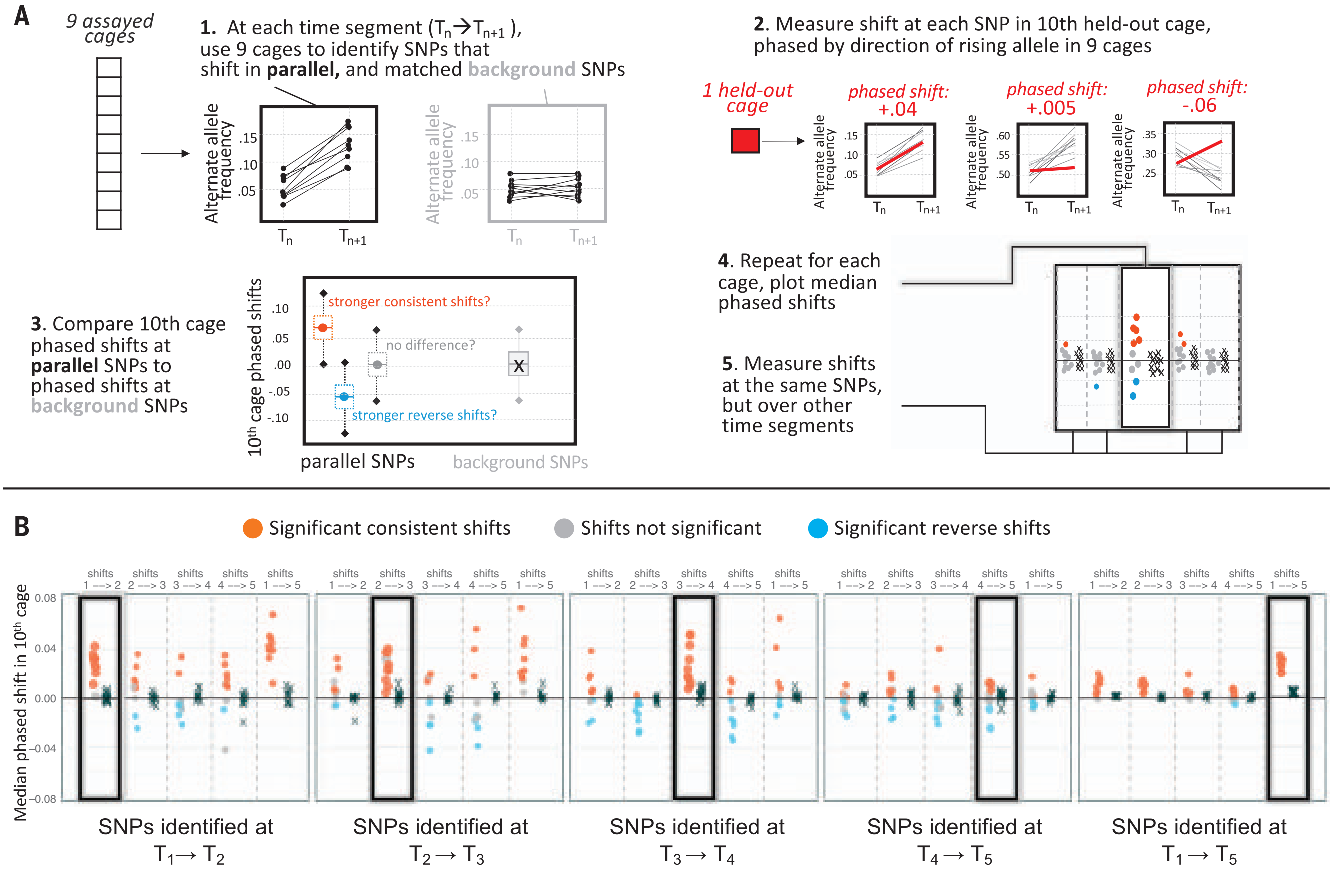


Fig. 4. Using genomic data to test for evolutionary parallelism indicative of adaptation. (**A**) Graphical description of the leave-one-out 10-fold cross-validation process for significant sites. In each round, significantly parallel sites (FDR < 0.05, effect size > 2%) at each time segment were identified using nine of the 10 cages, then the shift at those sites in the 10th left-out cage was measured, after phasing such that positive values represent shifts in the same direction as the nine assayed cages and negative values represent shifts in the reverse direction. The set of phased shifts at parallel sites was compared to phased shifts at background sites matched for chromosome and initial frequency and assigned to one of three significance bins: consistent (orange) or reverse (blue), or no significant difference from background (gray). Shifts at these same sites over other time segments were also measured, phased, and assigned to significance bins. (**B**) The median shift for each set of parallel sites (circles) and background sites (x marks) is plotted for each left-out cage. Each block of five panels represents shifts at the same sets of sites, those identified as parallel at the time segment labeled below the block. Shifts measured at that same time segment are highlighted in the panel with a dark shadowed outline.

and distinct signatures of adaptation over each interval and over the whole duration of the experiment. Hence, inferred rates of adaptation are dependent on the time scale of sampling (*13*). Our results clearly illustrate the extent to which lower-resolution temporal sampling would obscure the inference of adaptive tracking. Although sites identified during individual time intervals often showed median shifts of >2% in a single month, the strongest parallel sites detected from lower-resolution sampling (i.e., sampling only at T_1 and T_5) showed smaller monotonic shifts at each interval (on average, 0.6% per month). Moreover, the magnitude of this discrepancy varied widely over time. Taken together, these results underscore the value of high-resolution temporal sampling in revealing the existence of both temporally variable and temporally consistent directional selective forces.

Identifying the genomic architecture and putative functions of causal loci

The number and genomic locations of causal loci involved in adaptation are central to understanding the mechanics of the adaptive process (*28*). To quantify the genomic architecture of adaptation, we examined the distribution of parallel sites across the genome and developed an algorithm to differentiate putatively independent targets of selection from the sites whose shifts could largely be ascribed to linkage disequilibrium and genomic hitchhiking (genetic draft). We first fit allele frequencies from all 10 cages to a GLM and identified significantly parallel sites (fig. S8) at each time segment (n = 4274) and across the whole experiment (n = 5036), yielding 9310 significant shifts overall (Fig. 5A and table S4) (*27*). As expected from the leave-one-out analysis, the sets were largely nonoverlapping: The 9310 detected parallel shifts occurred at 9000 unique SNPs. Moreover, at each time interval and across the whole experiment, parallel sites were strongly clustered (empirical $P < 0.01$; fig. S9) and also showed significantly higher average linkage values than the matched control sites ($P < 10^{-16}$, paired t test; fig. S10) (*27*), which suggests that most parallel sites were merely linked to causal loci rather than being causal themselves.

We next identified the minimum number of independent genetic loci under selection using an algorithm that aggregated the parallel sites into clusters of linked sites (fig. S10) (*27*). This algorithm clustered 8214 parallel SNPs detected across all the time segments [~90% of all SNPs at a false discovery rate (FDR) of <0.05] into 165 unlinked independent clusters (Fig. 5A and table S5). These clusters were found on every chromosome and at every time segment, with an average of 4.5 clusters per chromosome per month. Simulations confirmed that although interference among multiple causal sites can temper shifts at any individual site, the number of clusters detected here still fell well within the realm of plausible selection landscapes. Specifically, when allele frequency trajectories for pairs, groups of 5, or groups of 10 selected loci were simulated simultaneously on the same chromosome, the majority (61.5%) of simulated selected sites required selection strengths no greater than s = 0.5 to achieve a minimum shift of 2% per monthly time segment, and the vast majority (80.2%) required selection strengths no greater than s = 1.

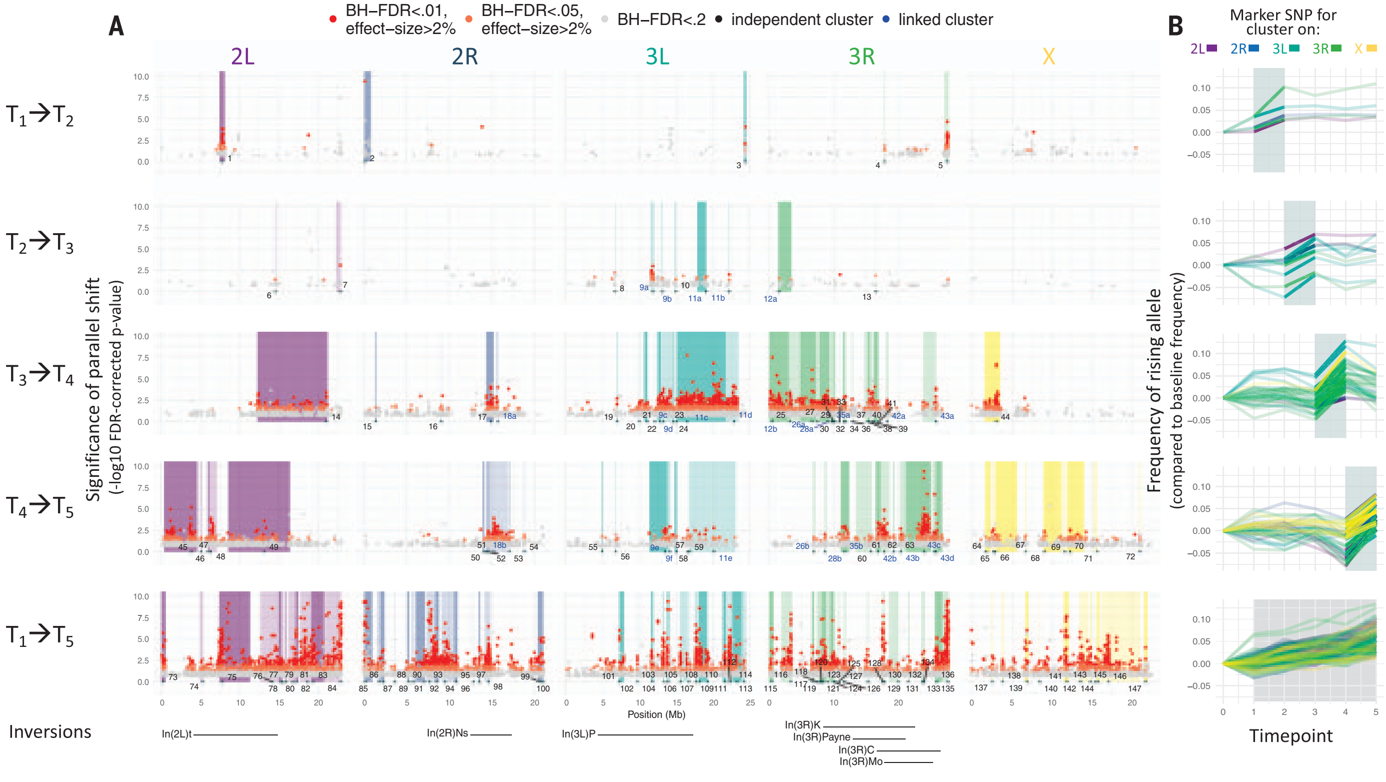


Fig. 5. The genomic architecture of parallel allele frequency change over time. (**A**) Manhattan plot of sites with significant parallel allele frequency shifts over time in 10 replicate cages. Each dot shows the $-\log_{10}$ of the FDR-corrected P value (y axis) corresponding to the significance of the allele frequency shift at a given SNP position (x axis) over a given time segment of the experiment (rows). Only SNPs with an FDR of <0.2 are shown, and dots are colored according to three significance bins (top). Shaded areas indicate regions of the genome that are likely driven by the same causal site, as defined by a clustering algorithm accounting for SNP linkage. Each clustered genome block is identified by a number marking the position of the top parallel SNP. Clusters from different time segments that are significantly linked ("superclusters") are given the same number, labeled in blue. The positions of seven common chromosomal inversions are indicated below. (**B**) Allele frequency trajectories are shown for the top marker SNP from each cluster. Each trajectory is translated to show allele frequency change relative to initial frequency in the baseline population and phased to show the frequency of the rising allele at the time segment in which the cluster was identified. The time segment over which the SNPs were identified as outliers is shaded in gray.

Furthermore, although inversions can drive patterns of adaptation in *Drosophila* (*29*, *30*), no inversion markers were found among the parallel sites, and only three of the 165 clusters were strongly linked to inversions with average $R^2 > 0.1$ (table S7 and fig. S11). Combining clusters from all time segments, 61% of all assayed SNPs and 62% of the genome were contained in at least one cluster, highlighting the pervasive impact of short-term adaptive evolution at tens to hundreds of independent selected sites on allele frequencies genome-wide.

The genomic distribution and frequency shifts of these clusters suggested rapid changes in the targets and direction of selection over time. Specifically, 36 of the 90 clusters (40%) identified at a specific monthly time interval did not overlap any clusters identified at other monthly intervals, which suggests that selection at these loci was limited to 1 month. Among the remaining 54 clusters, only 27 (50%) contained SNPs that were significantly linked to SNPs in the cluster they overlapped. These 27 clusters formed nine distinct "superclusters" (Fig. 5A) with high internal linkage, representing genomic regions in which allele frequencies shifted significantly in multiple monthly intervals. Strikingly, in five of the six superclusters involving a cluster from $T_3 \rightarrow T_4$ linked to a cluster from $T_4 \rightarrow T_5$, 90% of SNPs flipped direction between months, and in the sixth cluster >80% flipped direction, together totaling 10,464 SNPs that flipped direction (fig. S12). A smaller majority of SNPs (67%) flipped in the supercluster formed by a cluster from $T_2 \rightarrow T_3$ linked to a cluster from $T_3 \rightarrow T_4$. Finally, in the two superclusters involving sets of linked clusters from three different time segments ($T_2 \rightarrow T_3$, $T_3 \rightarrow T_4$, $T_4 \rightarrow T_5$), together covering more than 5 Mb of chromosome arm 3L, most SNPs (72% and 85%, respectively) flipped direction twice. We further confirmed that similar dynamics characterized the full set of putatively causal SNPs by choosing the SNP with the strongest parallelism *P* value in each cluster and examining its trajectory (Fig. 5B). Even though the initial frequencies of these marker SNPs (fig. S13) and the exact shape of their trajectories varied widely, we observed a consistent trend: Markers for the clusters identified at an individual monthly time interval often changed little during other months or even moved in the opposite direction (especially clusters identified at $T_3 \rightarrow T_4$), whereas markers for clusters identified across the whole experiment tended to shift evenly and monotonically over time. The analysis of overlapping clusters and marker SNPs revealed patterns similar to those seen in individual SNP-based analyses, together supporting an oligogenic and rapid adaptive response to momentary selection pressures that often results in strong and rapidly fluctuating selection.

We next tested whether the identified genomic targets of this rapidly fluctuating selection were associated with any specific phenotypic traits (*31*) or pathways. We specifically investigated the set of 111 genes—one per cluster—that overlapped with the cluster's top marker SNP. This set of genes is strongly enriched ($P < 0.001$ in all cases) for genes with a known phenotypic effect (85 genes), and more specifically for genes involved in behavior (27 genes), cell-to-cell signaling (34 genes), neuronal function (25 genes), synaptic function (14 genes), and the central nervous system (21 genes) (table S6). Many of these genes are crucial to core developmental and signaling pathways including the Wnt signaling pathway [genes *frizzled2* (the receptor of *wingless*), *armadillo* (β-catenin), *sgg* (GSK3), *flo2* (long-range Wnt signaling), *reck* (regulation of Wnt signaling), *huwe1* (negative regulation of Wnt signaling), and *tkv* (dpp/BMP signaling)]. Strikingly, one cluster marker SNP is found in *SNF4A*γ, the gamma subunit of the central metabolic switch kinase AMPK (adenosine 5′-monophosphate–activated protein kinase). *SNF4A*γ is one of two key genes previously found to be involved in adaptation to high temperature during experimental evolution of a sibling species, *D. simulans* (*32*). On balance, these patterns suggest that adaptive tracking in our outdoor mesocosms may be driven by shifts in the function of neuronal, metabolic, and development pathways that modulate sensing and regulatory processes and affect the way environmental cues are interpreted by the organism.

Discussion

The phenotypic and genomic patterns we observed are consistent with a form of adaptive tracking in which (i) populations adapt in response to continuous environmental shifts, (ii) parallel evolution is driven by strong selection on multiple phenotypes and on a substantial number (tens to hundreds) of strongly selected genetic variants, (iii) the identity of the phenotypes and variants under selection changes considerably over short time scales, and (iv) selection operates at multiple time scales, acting in a consistent direction across the whole experiment on some variants and phenotypes but rapidly fluctuating in direction and magnitude at others (*33*). This fluctuating selection leads to inferred rates of adaptation being slower when measured from the beginning to the end of the experiment as compared to single monthly intervals. The observed pattern that evolutionary rates are fastest when measured over shorter time scales may be driven by fluctuating selection (*13*, *34*).

The pace, complex architecture of adaptation, and temporal evolution of some phenotypes in our field cages are generally consistent with prior observations of seasonal evolution in natural temperate populations of *D. melanogaster* (*21*, *35*–*37*). However, with additional temporal resolution and replication, we detected rapidly fluctuating patterns of adaptation that suggest that populations of *D. melanogaster* are continuously and adaptively tracking the environment; this is surprising but not implausible given the pace of environmental change (*38*). These patterns also imply that segregating functional variation is abundant and that much of the segregating variation in fitness is likely due to balancing selection (*39*), including temporally fluctuating selection that maintains genetic variation (*14*, *40*, *41*). The functional analysis of the genomic regions under selection further suggests that the rapid adaptation detected here is likely driven by modulation of high-level signaling pathways that feed into developmental and neuronal functions capable of modifying multiple phenotypes in a coordinated fashion. This may explain how selection can rapidly modify so many ostensibly unrelated phenotypes at the same time.

Our experiment shows that it is possible to observe adaptive tracking in real time, thereby providing a new lens to study the synchronous ecological and evolutionary dynamics of natural populations. We have focused here on *D. melanogaster*, but the environmental and organismal features that gave rise to adaptive tracking, such as the presence of strongly shifting environmental pressures on generational time scales, are likely common (*7*, *8*, *42*, *43*). Understanding the complex interplay among environmental change, population dynamics, standing genetic variation, and trait architecture that dictates the extent of adaptive tracking is a considerable challenge. Determining whether adaptive tracking is a general feature of natural populations and defining the factors that shape the extent of adaptive tracking could be transformative in understanding the generation and maintenance of biodiversity.

REFERENCES AND NOTES

1. L. B. Slobodkin, *Growth and Regulation of Animal Populations* (Dover, 1980).
2. A. Charmantier *et al.*, Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science* **320**, 800–803 (2008). doi: 10.1126/science.1157174; pmid: 18467590
3. A. M. Simons, Modes of response to environmental change and the elusive empirical evidence for bet hedging. *Proc. R. Soc. B* **278**, 1601–1609 (2011). doi: 10.1098/rspb.2011.0176; pmid: 21411456
4. D. N. Reznick, F. H. Shaw, F. H. Rodd, R. G. Shaw, Evaluation of the rate of evolution in natural populations of guppies (*Poecilia reticulata*). *Science* **275**, 1934–1937 (1997). doi: 10.1126/science.275.5308.1934; pmid: 9072971
5. P. R. Grant, B. R. Grant, Unpredictable evolution in a 30-year study of Darwin's finches. *Science* **296**, 707–711 (2002). doi: 10.1126/science.1070315; pmid: 11976447
6. N. G. Hairston Jr., S. P. Ellner, M. A. Geber, T. Yoshida, J. A. Fox, Rapid evolution and the convergence of ecological and evolutionary time. *Ecol. Lett.* **8**, 1114–1127 (2005). doi: 10.1111/j.1461-0248.2005.00812.x

7. R. D. H. Barrett, S. M. Rogers, D. Schluter, Natural selection on a major armor gene in threespine stickleback. *Science* **322**, 255–257 (2008). doi: 10.1126/science.1159978; pmid: 18755942
8. R. D. H. Barrett *et al*., Linking a mutation to survival in wild mice. *Science* **363**, 499–504 (2019). doi: 10.1126/science.aav3824; pmid: 30705186
9. D. J. Rennison, S. M. Rudman, D. Schluter, Genetics of adaptation: Experimental test of a biotic mechanism driving divergence in traits and genes. *Evol. Lett.* **3**, 513–520 (2019). doi: 10.1002/evl3.135; pmid: 31636943
10. S. M. Rudman *et al*., Microbiome composition shapes rapid genomic adaptation of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 20025–20032 (2019). doi: 10.1073/pnas.1907787116; pmid: 31527278
11. J. B. S. Haldane, The cost of natural selection. *J. Genet.* **55**, 511–524 (1957). doi: 10.1007/BF02984069
12. E. A. Boyle, Y. I. Li, J. K. Pritchard, An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186 (2017). doi: 10.1016/j.cell.2017.05.038; pmid: 28622505
13. P. D. Gingerich, Rates of evolution: Effects of time and temporal scaling. *Science* **222**, 159–161 (1983). doi: 10.1126/science.222.4620.159; pmid: 17741657
14. M. Turelli, N. H. Barton, Polygenic variation maintained by balancing selection: Pleiotropy, sex-dependent allelic effects and G x E interactions. *Genetics* **166**, 1053–1079 (2004). doi: 10.1093/genetics/166.2.1053; pmid: 15020487
15. R. C. Lewontin, *The Genetic Basis of Evolutionary Change* (Columbia Univ. Press, 1974).
16. M. T. Kinnison, N. G. Hairston Jr., A. P. Hendry, Cryptic eco-evolutionary dynamics. *Ann. N.Y. Acad. Sci.* **1360**, 120–144 (2015). doi: 10.1111/nyas.12974; pmid: 26619300
17. D. Schluter, L. Nagel, Parallel speciation by natural selection. *Am. Nat.* **146**, 292–301 (1995). doi: 10.1086/285799
18. S. Rajpurohit *et al*., Spatiotemporal dynamics and genome-wide association analysis of desiccation tolerance in Drosophila melanogaster. *Mol. Ecol.* **27**, 3525–3540 (2018). doi: 10.1111/mec.14814; pmid: 30051644
19. T. N. Grainger, S. M. Rudman, P. Schmidt, J. M. Levine, Competitive history shapes rapid evolution in a seasonal climate. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2015772118 (2021). doi: 10.1073/pnas.2015772118; pmid: 33536336
20. S. Tilk *et al*., Accurate allele frequencies from ultra-low coverage pool-seq samples in evolve-and-resequence experiments. *G3* **9**, 4159–4168 (2019). doi: 10.1534/g3.119.400755; pmid: 31636085
21. E. L. Behrman, S. S. Watson, K. R. O'Brien, M. S. Heschel, P. S. Schmidt, Seasonal variation in life history traits in two Drosophila species. *J. Evol. Biol.* **28**, 1691–1704 (2015). doi: 10.1111/jeb.12690; pmid: 26174167
22. T. Flatt, Life-History Evolution and the Genetics of Fitness Components in *Drosophila melanogaster*. *Genetics* **214**, 3–48 (2020). doi: 10.1534/genetics.119.300160; pmid: 31907300
23. J. B. S. Haldane, Suggestions as to quantitative measurement of rates of evolution. *Evolution* **3**, 51–56 (1949). doi: 10.1111/j.1558-5646.1949.tb00004.x; pmid: 18115117
24. P. D. Gingerich, Quantification and comparison of evolutionary rates. *Am. J. Sci.* **293A**, 453–478 (1993). doi: 10.2475/ajs.293.A.453
25. A. P. Hendry, T. J. Farrugia, M. T. Kinnison, Human influences on rates of phenotypic change in wild animal populations. *Mol. Ecol.* **17**, 20–29 (2008). doi: 10.1111/j.1365-294X.2007.03428.x; pmid: 18173498
26. T. F. C. Mackay *et al*., The Drosophila melanogaster Genetic Reference Panel. *Nature* **482**, 173–178 (2012). doi: 10.1038/nature10811; pmid: 22318601
27. See supplementary materials.
28. J. Stapley *et al*., Adaptation genomics: The next generation. *Trends Ecol. Evol.* **25**, 705–712 (2010). doi: 10.1016/j.tree.2010.09.002; pmid: 20952088
29. T. Dobzhansky, Genetics of natural populations XVI: Altitudinal and seasonal changes produced by natural selection in certain populations of Drosophila pseudoobscura and Drosophila persimilis. *Genetics* **33**, 158–176 (1948). doi: 10.1093/genetics/33.2.158; pmid: 18856563
30. M. Kapun, T. Flatt, The adaptive significance of chromosomal inversion polymorphisms in Drosophila melanogaster. *Mol. Ecol.* **28**, 1263–1282 (2019). doi: 10.1111/mec.14871; pmid: 30230076
31. M.-P. Weng, B.-Y. Liao, modPhEA: Model organism Phenotype Enrichment Analysis of eukaryotic gene sets. *Bioinformatics* **33**, 3505–3507 (2017). doi: 10.1093/bioinformatics/btx426
32. F. Mallard, V. Nolte, R. Tobler, M. Kapun, C. Schlötterer, A simple genetic basis of adaptation to a novel thermal environment results in complex metabolic rewiring in Drosophila. *Genome Biol.* **19**, 119 (2018). doi: 10.1186/s13059-018-1503-4; pmid: 30122150
33. G. Bell, Fluctuating selection: The perpetual renewal of adaptation in variable environments. *Philos. Trans. R. Soc. London Ser. B* **365**, 87–97 (2010). doi: 10.1098/rstb.2009.0150; pmid: 20008388
34. A. P. Hendry, M. T. Kinnison, The pace of modern life: Measuring rates of contemporary microevolution. *Evolution* **53**, 1637–1653 (1999). doi: 10.1111/j.1558-5646.1999.tb04550.x; pmid: 28565449
35. P. S. Schmidt, D. R. Conde, Environmental heterogeneity and the maintenance of genetic variation for reproductive diapause in Drosophila melanogaster. *Evolution* **60**, 1602–1611 (2006). doi: 10.1111/j.0014-3820.2006.tb00505.x; pmid: 17017061
36. A. O. Bergland, E. L. Behrman, K. R. O'Brien, P. S. Schmidt, D. A. Petrov, Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in Drosophila. *PLOS Genet.* **10**, e1004775 (2014). doi: 10.1371/journal.pgen.1004775; pmid: 25375361
37. H. E. Machado *et al*., Broad geographic sampling reveals the shared basis and environmental correlates of seasonal adaptation in *Drosophila*. *eLife* **10**, e67577 (2021). doi: 10.7554/eLife.67577; pmid: 34155971
38. C. A. Botero, F. J. Weissing, J. Wright, D. R. Rubenstein, Evolutionary tipping points in the capacity to adapt to environmental change. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 184–189 (2015). doi: 10.1073/pnas.1408589111; pmid: 25422451
39. B. Charlesworth, Causes of natural variation in fitness: Evidence from studies of Drosophila populations. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 1662–1669 (2015). doi: 10.1073/pnas.1423275112; pmid: 25572964
40. M. J. Wittmann, A. O. Bergland, M. W. Feldman, P. S. Schmidt, D. A. Petrov, Seasonally fluctuating selection can maintain polymorphism at many loci via segregation lift. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E9932–E9941 (2017). doi: 10.1073/pnas.1702994114; pmid: 29087300
41. J. Bertram, J. Masel, Different mechanisms drive the maintenance of polymorphism at loci subject to strong versus weak fluctuating selection. *Evolution* **73**, 883–896 (2019). doi: 10.1111/evo.13719; pmid: 30883731
42. A. Garcia-Elfring *et al*., Using seasonal genomic changes to understand historical adaptation to new environments: Parallel selection on stickleback in highly-variable estuaries. *Mol. Ecol.* **30**, 2054–2064 (2021). doi: 10.1111/mec.15879; pmid: 33713378
43. M. A. Ehrlich, D. N. Wagner, M. F. Oleksiak, D. L. Crawford, Polygenic selection within a single generation leads to subtle divergence among ecological niches. *Genome Biol. Evol.* **13**, evaa257 (2021). doi: 10.1093/gbe/evaa257; pmid: 33313716
44. S. I. Greenblum, Dros-adaptive-tracking: v.1.0. Zenodo (2022); doi: 10.5281/zenodo.5839005.

ACKNOWLEDGMENTS

We thank A. Berry, M. Exposito-Alonso, H. Fraser, D. Hartl, J. Levine, E. Mordecai, D. Schluter, members of the Kelley and Cornejo labs, the Porter lab, the King lab, the Petrov lab, the Schmidt lab, and two anonymous reviewers for helpful comments and discussions. **Funding:** Supported by NIH grants F32GM122216 (S.I.G.), R01GM100366 and R01GM137430 (P.S.), and R35GM118165 and R01GM10036601 (D.A.P.). **Author contributions:** P.S. designed research. S.R., N.J.B., J.H., and P.S. conducted the experiment. T.Y. and S.T. prepared sequencing libraries. S.M.R., S.I.G., D.A.P., and P.S. analyzed data and wrote and revised the manuscript. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** We have deposited census and phenotypic evolution data in the Dryad digital repository (doi: 10.5061/dryad.9zw3r22gg). Sequenced founder lines can be found at NCBI Accession PRJNA722305. Sequencing data from evolved cages can be found at NCBI Accession PRJNA793595. Scripts for the genomic analysis and simulations can be found at https://doi.org/10.5281/zenodo.5839005 (*44*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj7484
Materials and Methods
Figs. S1 to S13
Tables S1 to S7
References (*45*–*54*)

2 June 2021; resubmitted 14 October 2021
Accepted 1 February 2022
10.1126/science.abj7484

Your Legacy to Science

AN ESTATE GIFT TO THE
AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

Since 1848, our founding year, the American Association for the Advancement of Science (AAAS) has been deeply committed to advancing science, engineering and innovation around the world for the benefit of all people.

By making AAAS a beneficiary of your will, trust, retirement plan or life insurance policy, you become a member of our 1848 Society, joining Thomas Edison, Alexander Graham Bell and the many distinguished individuals whose vision led to the creation of AAAS and our world-renowned journal, *Science*, so many years ago.

Unlike many of its peers, *Science* is not for-profit. Your estate gift would provide long-term financial stability and durable annual income that will support operations and competitive innovation for years to come. **This support is vital.**

"As a teacher and instructor, I bear responsibility for the younger generations. If you have extra resources, concentrate them on organizations, like AAAS, that are doing work for all."

—Prof. Elisabeth Ervin-Blankenheim, 1848 Society member

If you intend to include AAAS in your estate plans, provide this information to your lawyer or financial adviser:

Legal Name: American Association for the Advancement of Science
Federal Tax ID Number: 53-0196568
Address: 1200 New York Avenue, NW, Washington, DC 20005

If you would like more information on making an estate gift to AAAS, cut out and return the form below or send an email to philanthropy@aaas.org. Additional details are also available online at **www.aaas.org/1848Society**.

1848
SOCIETY
AAAS

cut here ✂

Yes, I would like more information about joining the AAAS 1848 Society.

PLEASE CONTACT ME AT:

Name: ____________________

Address: ____________________

City: ____________ State: ______ Zip code: ____________ Country: ______

Email: ____________________ Phone: ____________

RETURN THIS FORM TO:
AAAS Office of Philanthropy and Strategic Partnerships • 1200 New York Avenue, NW • Washington, DC 20005 USA

RESEARCH ARTICLES

HUMAN GENOMICS

Multiple causal variants underlie genetic associations in humans

Nathan S. Abell[1]*, **Marianne K. DeGorter**[2], **Michael J. Gloudemans**[3], **Emily Greenwald**[1], **Kevin S. Smith**[2], **Zihuai He**[4,5], **Stephen B. Montgomery**[1,2]*

Associations between genetic variation and traits are often in noncoding regions with strong linkage disequilibrium (LD), where a single causal variant is assumed to underlie the association. We applied a massively parallel reporter assay (MPRA) to functionally evaluate genetic variants in high, local LD for independent cis-expression quantitative trait loci (eQTL). We found that 17.7% of eQTLs exhibit more than one major allelic effect in tight LD. The detected regulatory variants were highly and specifically enriched for activating chromatin structures and allelic transcription factor binding. Integration of MPRA profiles with eQTL/complex trait colocalizations across 114 human traits and diseases identified causal variant sets demonstrating how genetic association signals can manifest through multiple, tightly linked causal variants.

Genome-wide association studies (GWASs) have emerged as an important tool with which to assess the effect of individual genetic variants on phenotypes, ranging from gene expression to complex traits and diseases (*1*, *2*). However, because of linkage disequilibrium (LD), it is challenging to identify a single causal variant among multiple correlated variants. To address this challenge, statistical and functional fine-mapping approaches have been developed to identify credible sets of variants that contain the causal variant (*3*). However, these approaches often cannot distinguish between proximal or highly linked variants and lack systematic prior information on the number of causal variants underlying association signals.

One approach to systematically identify causal variants while controlling for LD is applying massively parallel reporter assays (MPRAs). MPRAs measure the effects of synthetic DNA libraries on the expression of a reporter gene, typically luciferase or green fluorescent protein (GFP), containing a 3′ untranslated region (UTR) barcode (*4*). Such assays have screened potential regulatory elements in diverse cellular contexts and also have applications in saturation mutagenesis or tiling along regulatory regions of interest (*5*–*7*).

Beyond tests of regulatory function, MPRAs have also been applied to assay the differential regulatory effects of genetic variants (*8*–*10*). However, existing studies have either targeted variants with the strongest trait associations and/or applied extensive prior filtering limiting resolution of linked causal variants (*8*, *9*, *11*, *12*). In the yeast *Saccharomyces cerevisiae*, quantitative trait loci (QTL) mapping has identified loci containing multiple causal variants in tight LD, suggesting that the same genetic architecture may also underlie many human traits (*13*, *14*).

Results

Functional fine-mapping of eQTL reproducibly identifies regulatory and allelic hits

We applied an MPRA to systematically characterize causal variants underneath multiple expression QTL (eQTL) and GWAS loci. We selected independent, common, and top-ranked eQTL across 744 eGenes identified in the CEU cohort (which comprises Utah residents of Northern and Western European ancestry). Each eQTL had a median of six lead associated variants (range of 1 to 472) in perfect LD. For each lead variant, we identified all additional variants with a correlation coefficient (r^2) ≥ 0.85 that were associated with the same gene, as well as a set of variants (n = 2114 non-eQTLs) that were not associated with any gene's expression. Our final library included 30,893 variants, with a median of 50 variants per eQTL (range of 2 to 2824) (Fig. 1A).

For each variant, we identified 150–base pair (bp) sequences (centered on the variant) and generated a MPRA library by random barcoding (Fig. 1B). For allelic pairs, the fragment lengths and surrounding sequence were held constant to allow measurement of allele-specific effects. For indels, fragment lengths between allelic pairs differed by less than 9 bp. Furthermore, in sequences with multiple variants, distinct oligonucleotides (oligos) were designed for each possible haplotype, resulting in an average of 3.19 oligos per variant. Overall, this resulted in an assay of 49,256 total allelic pairs. After reporter gene insertion, the library was transfected into lymphoblastoid cell lines (LCLs) in triplicate, sequenced, and then quantified for each oligo.

To measure regulatory effects from oligo counts, we used negative binomial regression. For each variant, we computed the allele-independent regulatory effects of an oligo ("expression" effects) and the difference in regulatory effects between reference and alternative allele-containing oligos ("allelic" effects). We detected 8502 expression effects and 1264 allelic effects across all tested variants.

We observed a modest increase in the total number of MPRA hits in eQTLs relative to non-eQTLs (27 versus 26% for expression hits and 9 versus 8% for allelic hits), reflecting the low proportion of eQTL variants overall that are expected to be causal (Fig. 1, C and D). We observed a larger increase in allelic effect sizes among hits that are also eQTL versus non-eQTL (fig. S1D). This was the case when comparing MPRA hits between eQTL and non-eQTL for both expression effects [Kolmogorov-Smirnov (K-S) $P = 1.704 \times 10^{-4}$) and allelic effects (K-S $P = 0.0116$). Taken together, we obtained for each eQTL gene (eGene), a profile of allele-independent and -dependent effects across all highly associated proximal variants (Fig. 1E).

By design, a subset of tested variants (n = 782) were previously identified as expression-modulating variants in (*8*). This overlapping subset was highly enriched for expression and allelic effects (Fig. 1, C and D). Further, we observed that 89.6% of allelic MPRA hits in both datasets were directionally concordant (fig. S2A). From these results, we constructed a concordant, high-confidence "MPRA-positive" variant set that contains 250 variants with expression effects and 120 with allelic effects (fig. S2, B and C).

Diverse transcription factor programs contribute at eQTL

The large number of MPRA expression effects enabled identification of transcription factors (TFs) that affect gene expression within eQTLs. We observed widespread positive enrichment of chromatin immunoprecipitation–sequencing (ChIP-seq) peaks for multiple TFs in MPRA expression effects (n = 160 total TFs). Moreover, applying a more stringent filter (adjusted $P \leq 5 \times 10^{-10}$) increases these enrichments in most TFs (Fig. 2A and table S5). Although enrichments vary across a broad range (1.2- to 17-fold), many enriched TFs are members of the same family and exhibit highly correlated genome-wide binding profiles. This demonstrates the wide range of regulatory element

---

[1]Department of Genetics, School of Medicine, Stanford University, Stanford, CA 94305, USA. [2]Department of Pathology, School of Medicine, Stanford University, Stanford, CA 94305, USA. [3]Biomedical Informatics Program, Stanford University, Stanford, CA 94305, USA. [4]Department of Neurology and Neurological Sciences, Stanford University, Stanford, CA 94305, USA. [5]Quantitative Sciences Unit, Department of Medicine, Stanford University, Stanford, CA 94305, USA.
*Corresponding author. Email: nsabell@stanford.edu (N.S.A.); smontgom@stanford.edu (S.B.M.)

effects captured in our assay and pinpoints specific TFs that drive the regulatory effects of genetic variation.

We next evaluated histone modifications and observed enrichments for activating histone modifications but not for repressive marks such as histone H3 lysine 36 trimethylation (H3K36me3) (Fig. 2B). We also observed the strongest enrichments in chromatin accessibility regions that were tissue invariant or specific to the stromal A (representing JDP2 and other AP-1 TF families), lymphoid, and erythroid/myeloid tissue clusters, demonstrating detection of cell-type information encoded in accessible chromatin (Fig. 2C).

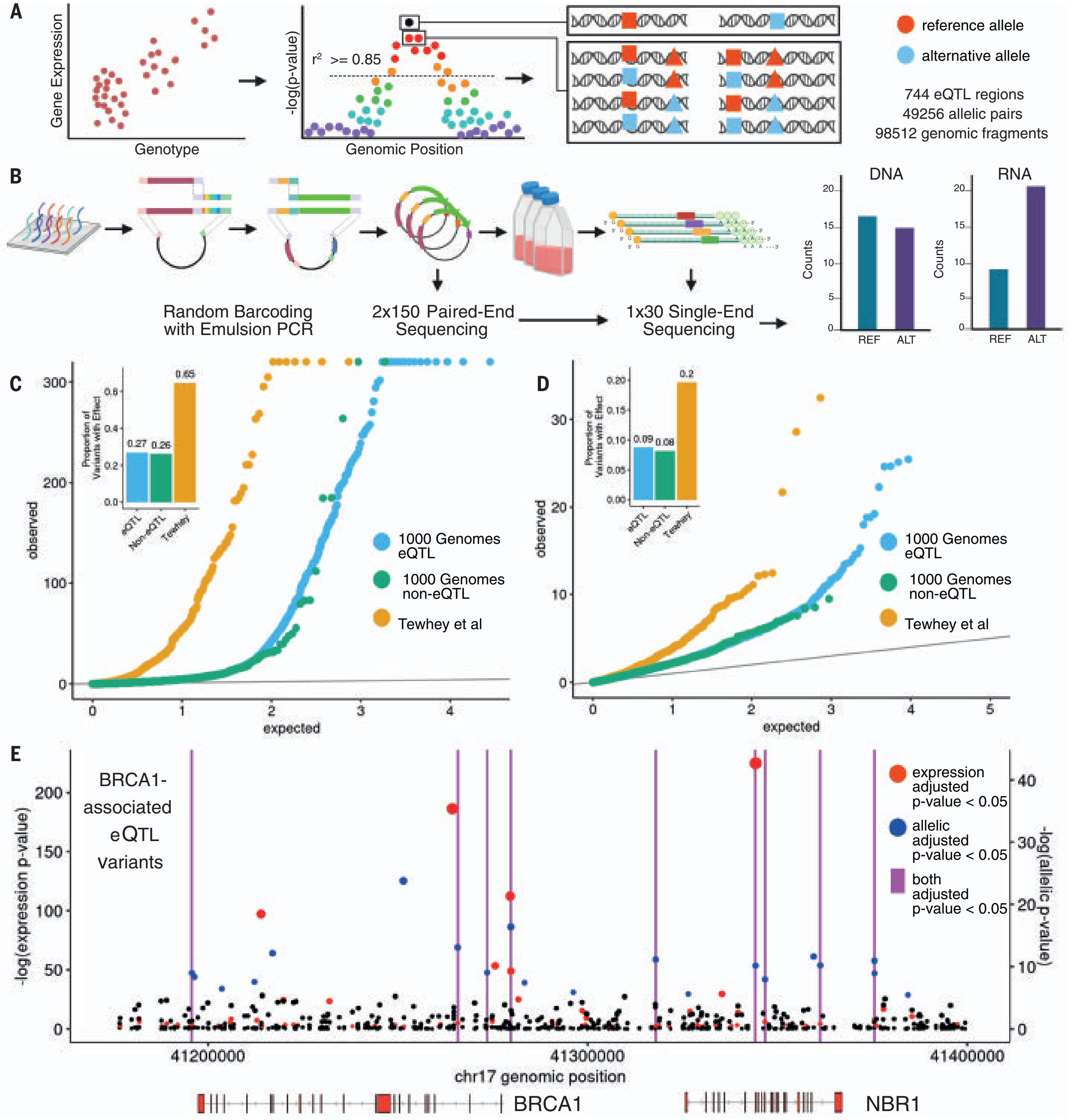


Fig. 1. Design and implementation of a variant-based MPRA. (**A**) Variant selection and oligonucleotide sequence design. (**B**) Random barcoding, sequencing, and expression of the MPRA library. (**C**) Distribution of eQTLs (orange) and non-eQTLs (blue) from the 1000 Genomes Project compared with Tewhey *et al.* (green) (8) variant expression *P* values (negative binomial regression) and relative effect proportions. (Inset) Proportion of tested variants that are significant MPRA hits. (**D**) Same as in (C) but with allelic *P* values (negative binomial regression). (**E**) Genomic position and unadjusted *P* values for all tested breast cancer 1 (BRCA1)–associated variants, with colors indicating Benjamini-Hochberg (BH) adjusted $P \leq 0.05$. Vertical magenta lines indicate positions of variants that are both expression and allelic MPRA hits.

Identifying regulatory variants through allelic transcription factor binding and chromatin accessibility

To identify specific TFs affected by regulatory variation, we first characterized whether the direction of allelic MPRA hits was concordant with single-nucleotide polymorphism evaluation by systematic evolution of ligands by exponential enrichment (SNP-SELEX) scores, a set of allele-specific binding models created from in vitro TF binding affinities. We observed global concordance (Fisher's exact $P = 3.43 \times 10^{-15}$) that was absent in other tested sites (Fisher's exact $P = 0.63$) (Fig. 2D).

Next, we computed the concordance proportion (how often a SNP-SELEX score for a specific TF was concordant with the MPRA allelic effect) for all TFs that overlapped at least three tested variants (Fig. 2E). The mean concordance proportion across TFs ($n = 59$) was 0.733 when using allelic MPRA hits and 0.505 when using other tested sites ($n = 91$) [binomial logistic generalized linear model (GLM) $P = 8.46 \times 10^{-12}$]. Although allelic MPRA hits were enriched in SNP-SELEX variants, only 13.5% of SNP-SELEX variants had an MPRA effect. This suggests that many allelic effects can be explained by altered TF binding, but altered binding itself does not typically affect transcription.

A similar pattern emerged when comparing allelic imbalance in accessible chromatin with MPRA allelic hits. We observed significant

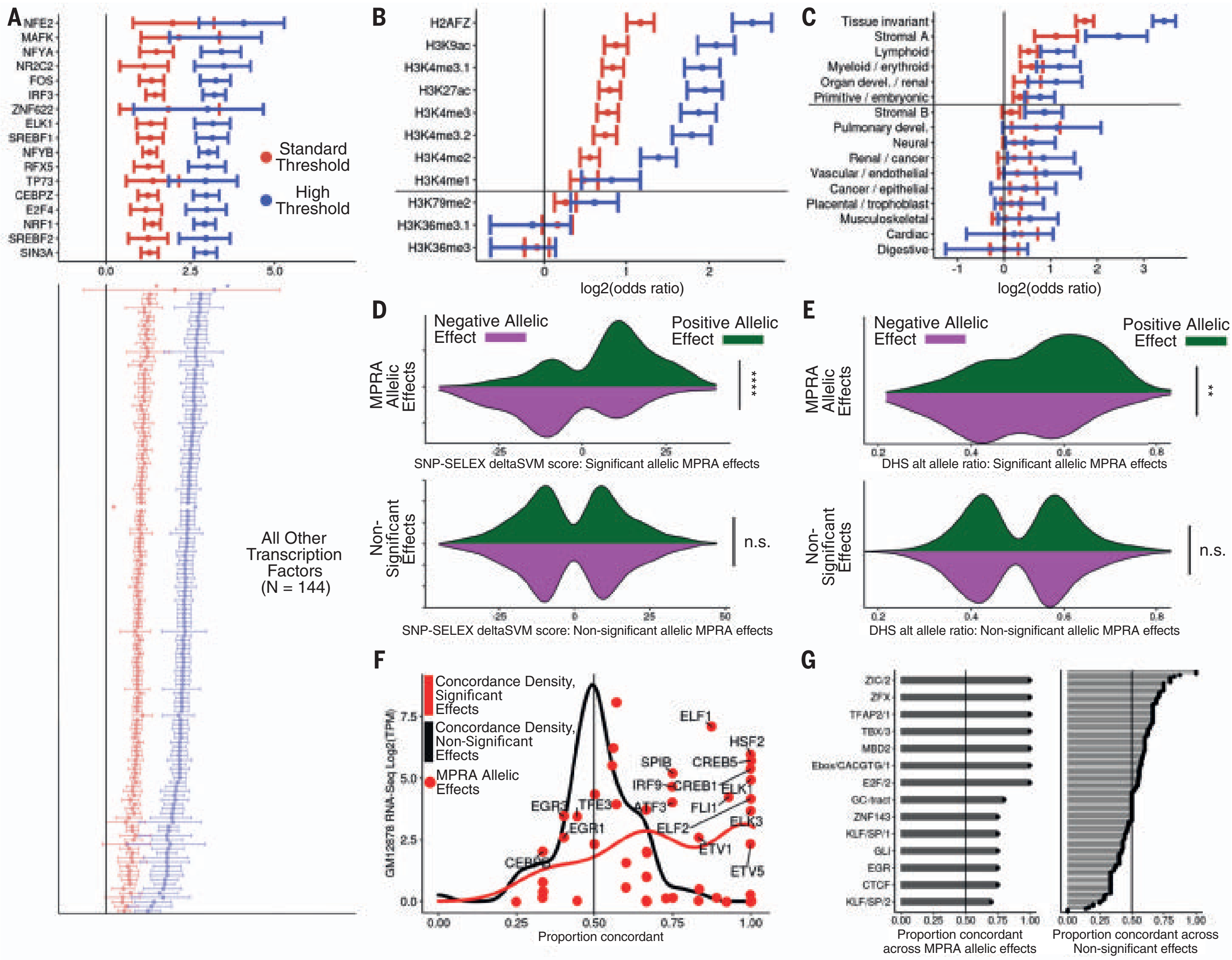


Fig. 2. General and allele-specific functional properties of regulatory variants. (**A**) Odds ratios and 95% confidence intervals for enrichment of peaks from 160 ENCODE ChIP-seq datasets within expression MPRA hits. Standard and high thresholds required an expression BH-adjusted $P < 5 \times 10^{-2}$ or $P < 5 \times 10^{-10}$, respectively. Only TFs with an enrichment adjusted $P < 0.005$ are shown, and listed TFs have BH-adjusted enrichment $P < 0.05$ and odds ratio > 5 (Fisher's exact test). (**B**) Same as in (A) but for histone modifications. Marks above the horizontal line have a BH-adjusted $P < 0.05$ at both thresholds. (**C**) Same as in (A) but for clustered chromatin accessibility regions in fragments with expression effects. (**D**) Distribution of SNP-SELEX deltaSVM scores at allele-specific binding variants stratified by MPRA allelic hit direction (color) and significance category (top and bottom). MPRA allelic hits have BH-adjusted expression and allelic $P \leq 0.05$, whereas nonsignificant variants have $P > 0.75$ (negative binomial regression). (**E**) Same as in (D) but for allelic imbalance in chromatin accessibility from ENCODE. (**F**) For all TFs evaluated in (D), comparison of the concordance proportion across MPRA variants with the expression of each included TF in GM12878 cells. Points indicate significant effect concordances. (**G**) Comparison of directional concordances within accessible chromatin motifs for (left) significant and (right) nonsignificant MPRA effects. Significance values for (C) and (D) were calculated with Fisher's exact test; $^{*}P < 0.05$, $^{**}P < 0.005$, $^{***}P < 0.0005$, $^{****}P < 5 \times 10^{-5}$.

concordance between allelic imbalance and MPRA allelic effect directions for allelic MPRA hits but not other variants (Fisher's exact $P = 7.33 \times 10^{-3}$ and $P = 0.839$, respectively) (Fig. 2F). Separation by functional footprints found within accessible chromatin regions revealed that several motifs, including Gli and a canonical E-box, were concordant across all allelic MPRA hits (Fig. 2G).

To further assess the relationship between regulatory variants and chromatin accessibility, we integrated chromatin accessible QTL (caQTL) data to identify variant annotations that increased MPRA signals. Using Encyclopedia of DNA Elements (ENCODE) allelic imbalance data, we separated all variants by whether they were inside or outside an associated peak. MPRA allelic hits were strongly concordant with allelic imbalance when inside their peaks but not when adjacent to them (Fisher's exact $P = 3.2 \times 10^{-5}$ and $P = 0.055$, respectively) (fig. S3A). Separately, in a set of caQTLs assessed across 10 population groups, variants that were caQTLs in multiple populations were more enriched in MPRA allelic hits than caQTLs shared in only a few populations (fig. S3B). Taken together, MPRA allelic hits were concordant with in vitro and in vivo measures of allelic regulatory activity, whereas other tested sites were directionally random.

MPRAs inform noncoding variant effect prediction

An ongoing challenge is to summarize and predict the regulatory effect of noncoding variants by using sequence and annotation alone. We evaluated whether genome-wide variant effect predictors could identify allelic MPRA hits. Using scores from Enformer, a neural network that predicts variant effects by incorporating sequence information, we observed significant enrichment of allelic MPRA hits in the top percentiles of Enformer scores (K-S test) (Fig. 3; A, inset, and B) (*15*). We next assessed all tested variants with their annotation principal components (aPCs) from FAVOR, an integrated variant effect prediction tool (*16*). We again observed enrichment of allelic MPRA hits for multiple aPCs. These enrichments were strongest for the TF and epigenetics-based aPCs, whereas others such as distance from transcription start site/transcription end site (TSS/TES) were similarly enriched in both allelic MPRA hits and all other tested sites (K-S test) (Fig. 3; C, inset, and D).

Both predictors could distinguish eQTL regions from genomic background; we also observed a positive enrichment in allelic MPRA hits up to the 50th percentile of these scores, with increasing enrichment at very high percentiles (Fig. 3, A and C). Despite this overlap, when comparing allelic MPRA hits with other tested variants, the distributions of all Enformer PCs and FAVOR aPCs, except for Distance-to-TSS/TES, were significantly different (K-S test) (Fig. 3, B and D). This suggests why functional fine-mapping approaches have not always benefitted from noncoding variant effect predictions while also showing that the highest genome-wide percentiles of these scores identify variants enriched for MPRA allelic effects.

Multiple causal regulatory variants in high LD underlie eQTL

To fine-map regulatory variants, we assessed MPRA hits within eQTLs. Across all loci, 76.7% (571 of 744) and 45.6% (339 of 744) had at least one expression or allelic MPRA hit, respectively (Fig. 4, A and B); 17.7% (132 of 744) had more than one allelic MPRA hit, indicating that an appreciable number of genetic associations contain multiple regulatory variants in high LD (Fig. 4B). Of allelic hits, 69% were in perfect LD in Europeans from the 1000 Genomes Project, which limited the use of statistical approaches (Fig. 4C). Even when additionally requiring a strong MPRA expression effect (|log 2 effect size| > 1.4), 6.3% of all eQTL contained multiple regulatory variants.

The degree to which eQTLs are composite products of multiple causal variants is unknown because of high LD. We assessed whether allelic MPRA hits found within eQTL were more likely to be concordant with eQTL effect direction than other tested sites. We found

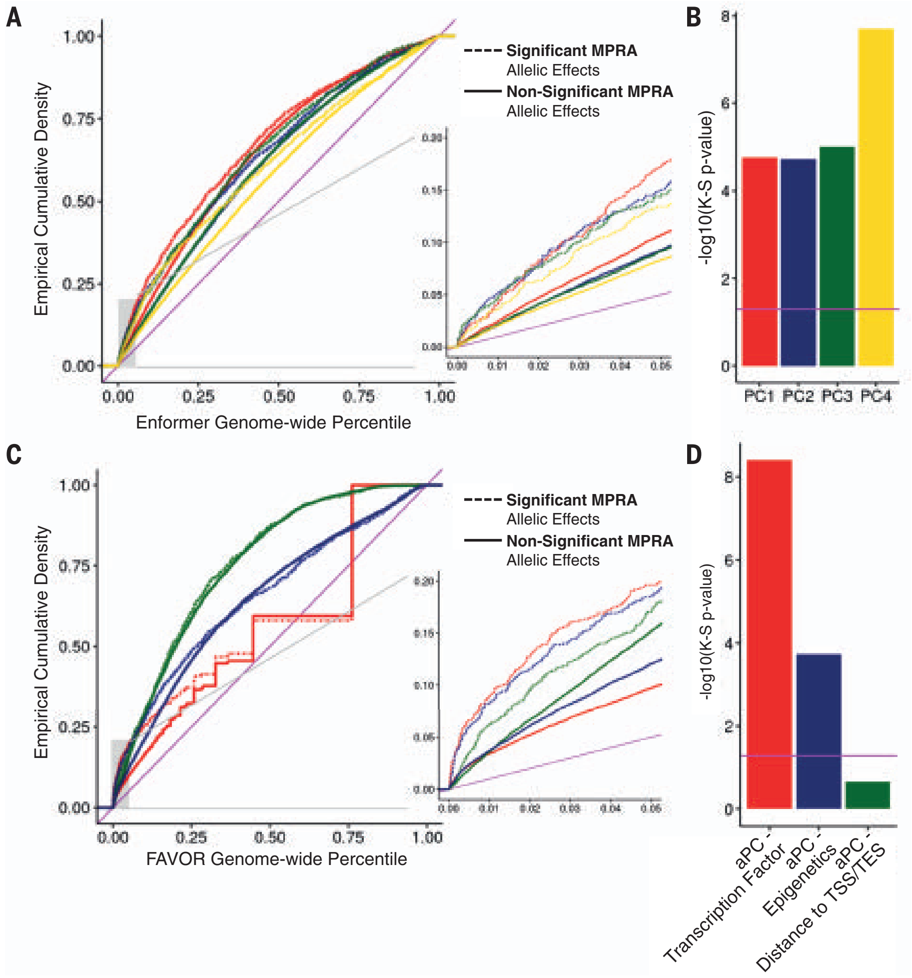


Fig. 3. Integrative noncoding variant effect prediction. (**A**) Empirical cumulative probability distribution of the first through fourth principal components (PC) scores from Enformer for allelic MPRA hits and other tested variants significant and nonsignificant MPRA allelic hits. Genome-wide percentiles were computed across all common variants in 1000 Genomes Phase 3. (Inset) A blow up of lower genome-wide percentile curves. (**B**) Significance of a K-S test comparing the empirical distributions of Enformer scores for significant and nonsignificant allelic MPRA hits v; magenta horizontal line indicates significance by K-S test ($P < 0.05$). (**C**) Same as in (A) except showing annotation principle components from FAVOR. Genome-wide percentiles were computed across all variants in TOPMed Freeze5. (**D**) Same as in (B), except testing FAVOR aPCs.

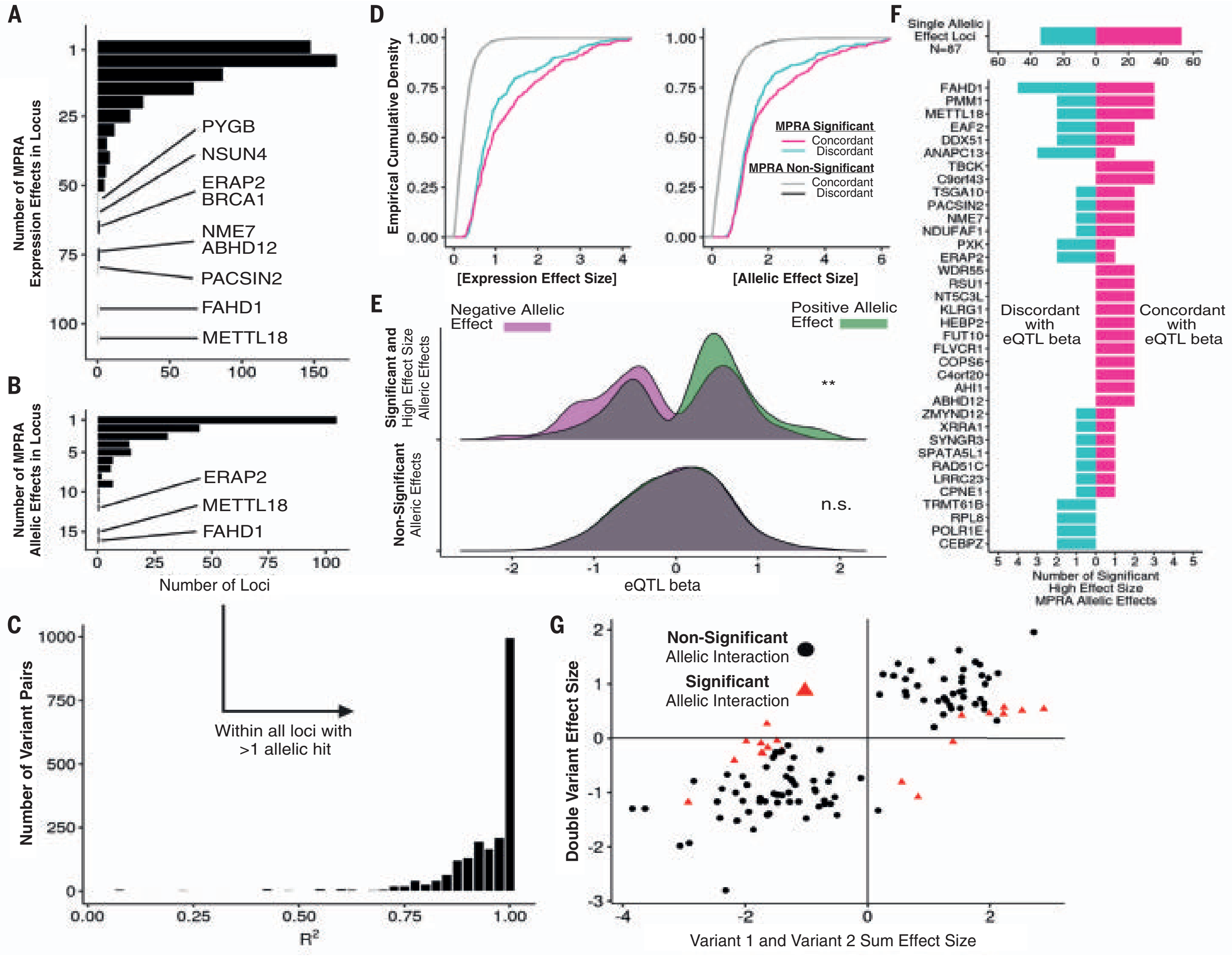


Fig. 4. Decomposition of allelic heterogeneity within regulatory loci. (**A**) Histograms of the number of expression MPRA hits per locus with BH-adjusted $P \leq 0.05$ (negative binomial regression). (**B**) Same as in (A) but requiring BH-adjusted $P \leq 0.05$ for allelic MPRA hits. (**C**) Distribution of LD r^2 values between all pairs of allelic MPRA hits within genes with multiple hits. (**D**) Cumulative distribution of effect sizes stratified by concordance. Concordance is defined as the sign of the allelic effect size matching the sign of eQTL beta. (**E**) Distribution of eQTL betas measured in GTEx v8 LCLs for strong MPRA hits (log expression effect size ≥ 1.4), stratified by MPRA allelic effect direction and significance from negative binomial regression. (**F**) Using the same variants as (E), counts of directionally concordant and discordant allelic MPRA hits across all loci. (**G**) Comparison of haplotype regression coefficients for variants tested individually or jointly; red points indicate allelic interaction BH-adjusted $P \leq 0.05$ (negative binomial regression). The x axis displays the sum of effect sizes associated with oligos containing each variant individually, and the y axis displays the effect size associated with the oligo containing both variants.

that expression and allelic MPRA effect sizes were larger for concordant variants compared with discordant variants (Fig. 4D). Across strong allelic MPRA hits ($|\log 2 \text{ effect size}| > 1.4$), we observed significant concordance with eQTL effect direction (Fisher's exact $P = 4.75 \times 10^{-3}$) (Fig. 4E) and the strongest examples of allelic heterogeneity (Fig. 4F).

To rule out study-specific effects, we verified that eQTL effect sizes were consistent across multiple studies (fig. S4A) (*17*, *18*). We found consistent patterns of concordance (fig. S4B). Additionally, to ensure that concordance patterns were not driven by individual eQTL with many concordant MPRA hits, we applied binomial count logistic regression to test whether concordance proportions were shifted between allelic MPRA hits and other tested sites. We found that allelic MPRA hits, but not other sites, were significantly concordant ($P = 2.85 \times 10^{-3}$) (fig. S4C). We further found that concordance persists through the top four ranked variants per eQTL, with the set of third-strongest MPRA hits across all eQTL having a concordance rate of 0.67 (fig S4D). Altogether, these results indicate that several eQTL regions contain multiple, concordant allelic MPRA hits.

Haplotype decomposition identifies allelic regulation that is unlikely to be observed by population sampling

A major advantage of synthetic library design is separation of extremely proximal variants that are unlikely to be naturally separated through recombination. Our library included 2097 pairs of eVariants within 75 bp. For these variants, we extended our statistical model to account for four haplotypes at each pair of variants and computed summary statistics for each of the three nonreference haplotypes (fig. S5A). We then selected all variants included in at least one haplotype allelic MPRA hit.

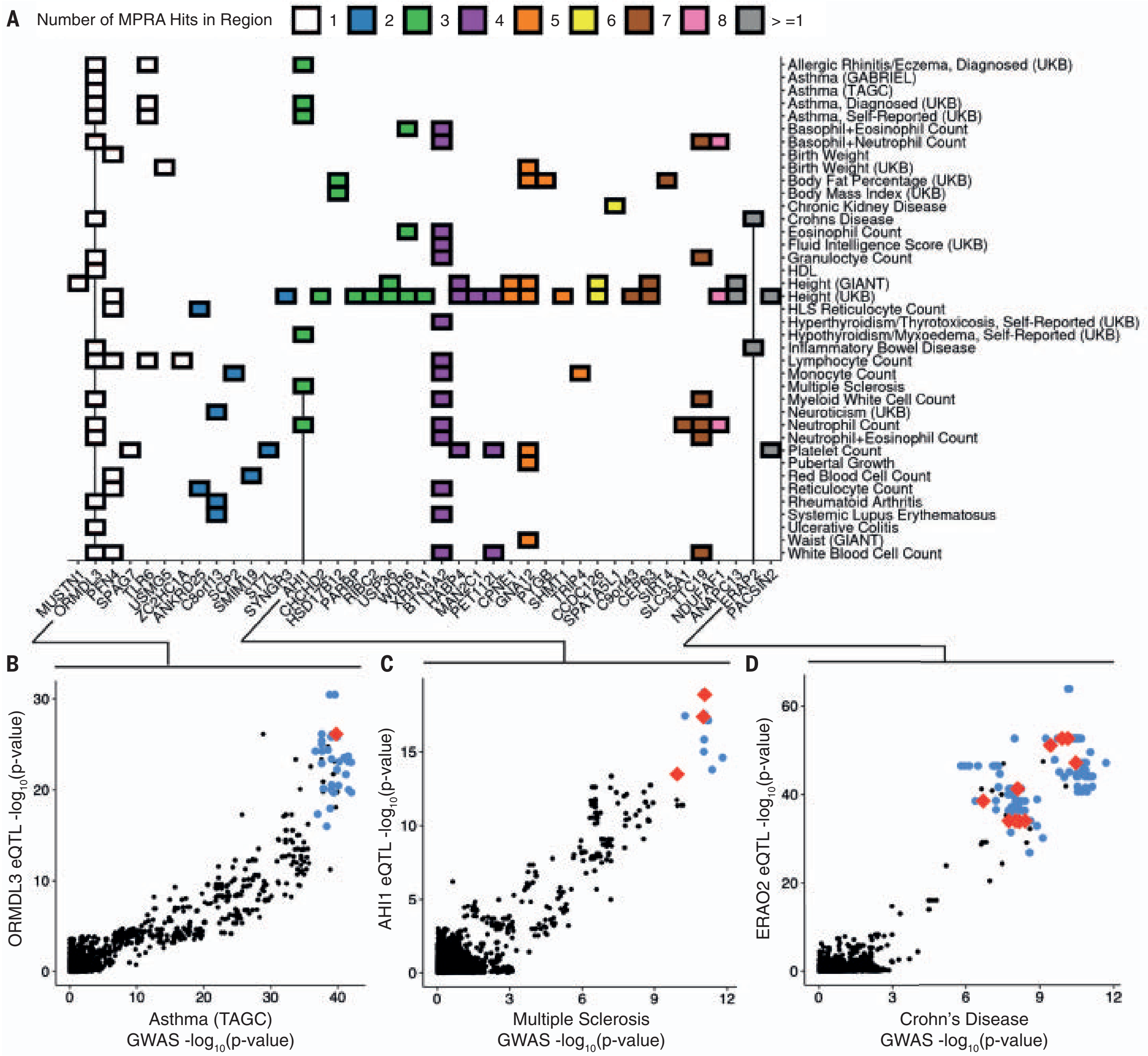


Fig. 5. Resolving complex trait associations with multiple causal variants. (**A**) Heatmap of significant colocalizations between eQTL loci and selected GWAS. Color indicates the number of allelic MPRA hits within the colocalized regions. (**B**) Comparison of genetic associations for asthma and *ORMDL3* expression in GTEx v8 LCLs. Red and blue points indicate allelic MPRA hits and other tested variants significant and nonsignificant allelic MPRA hits, respectively; black points indicate untested variants not included in our library. (**C**) Same as in (B) for associations with multiple sclerosis and *AHI1* expression. (**D**) Same as in (B) and (C) for associations with Crohn's disease and *ERAP2* expression, which was also colocalized with inflammatory bowel disease (fig. S5A). All GWAS and eQTL colocalizations are retrieved from (*22*), and lead variants were required to be genome-wide significant (reported GWAS $P \leq 5 \times 10^{-8}$ and reported eQTL $P \leq 5 \times 10^{-5}$) even if colocalization probability was high.

Combined, we identified 120 variant pairs (6.15% of all tests) with at least one haplotype allelic MPRA hit relative to all-reference sequence (negative binomial adjusted $P < 0.05$).

Most of the haplotype effects appeared additive, with a small number displaying nonadditivity (Fig. 4G). Our linear contrast test allowed us to identify these nonadditive interactions between allelic hits (fig. S5, B to D, and table S7). Of the variant pairs with at least one haplotype hit, 19 pairs also had a significant haplotype interaction effect (negative binomial adjusted $P < 0.05$; 14.7% of all significant haplotype effects and 0.91% of all tested variant pairs). Significant interactions were weaker than additive effects (fig. S5E) and rarely reversed the direction of individual allelic effects. These results support other studies that have identified nonadditive regulatory effects (*14*, *19*–*21*) and find that 14.7% of haplotype effects (only 0.91% of all tests) have evidence of nonadditivity.

Experimental fine-mapping of complex trait associations

To identify loci with shared genetic architecture between eQTL and human traits, we retrieved all genes tested in our dataset that had both an allelic MPRA hit and at least one LCL eQTL/GWAS colocalization (*22*). Out of 744 eGenes, 5.51% colocalized with at least one trait and contained at least one allelic MPRA hit. Most colocalizations contained more than one allelic MPRA hit (71.9% of colocalizations and 82.9% of eGenes), with some loci containing as many as 13 (Fig. 5A). This suggests that the default assumption of one causal variant, often used in fine mapping or GWAS colocalization, does not reflect causal variant biology at many regulatory regions. Traits with high-confidence colocalization were diverse, including blood-cell traits such as *ZC2HC1A*/lymphocyte count or *PACSIN2*/platelet count and highly polygenic traits such as *GNA12*/height (fig. S6).

The 17q21 locus contains the most extensively replicated genetic association with asthma, which colocalizes with *ORMDL3* eQTLs (Fig. 5B). This region contains a haplotype block with dozens of linked variants, flanked by two variants (rs4065275 and rs12936231) that induce loss and gain of CCCTC-binding factor (CTCF) binding, respectively. Further, other variants located between these two variants display allele-specific chromatin accessibility, histone modification, and CpG methylation (*23*, *24*). Altogether, the risk haplotype results in increased *ORMDL3* expression, which in turn negatively regulates interleukin-2 production in CD4$^+$ T cells. We identified a single allelic MPRA hit, rs12950743, that is linked to and located between the two CTCF variants (Fig. 5B). When tested by means of luciferase assay, this variant displayed a nominally significant but weak effect in the same direction as the MPRA (luciferase unpaired *t* test P = 0.035) (fig. S7A). Taken together, this suggests that two variants on the risk haplotype alter *CTCF* binding, leading to distinct regulatory contacts with their own allelic specificity.

By contrast, a different colocalization that included three active variants was *AHI1*, a well-characterized gene strongly associated with multiple sclerosis (MS) (Fig. 5C) (*25*). This region contains a strong eQTL and colocalization signal in LCLs; however, its causal variant(s) are unknown. We identified rs6908428, rs9399148, and rs761357 as allelic MPRA hits. We validated the allelic effects of these variants by means of luciferase assay and found that rs6908428 (luciferase unpaired *t* test $P = 5.1 \times 10^{-6}$) and rs761357 (unpaired *t* test $P = 7.6 \times 10^{-3}$) showed allelic differences consistent with the MPRA, whereas rs9399148 (unpaired *t* test P = 0.18) did not (fig. S7B). The first two of these variants have been highlighted by annotation overlap in prior studies of the role of *AHI1* in MS pathology, particularly interferon-γ production and CD4$^+$ T cell differentiation, but were severely limited by linkage across the risk haplotype (*25*). When screened against known TF binding motifs, we found that rs6908428 and rs761357 overlapped predicted binding motifs for SMAD3/4 and HNF1A, respectively. Unlike HNF1A, SMAD3/4 are expressed in LCLs, suggesting that rs6908428 may function to create a SMAD3/4 binding site (fig. S8A).

A complex multivariant colocalization was identified at *ERAP2*, an aminopeptidase functionally implicated in both inflammatory bowel disease and Crohn's disease (*26*). We detected 13 active variants that span a strongly linked haplotype. Although both eQTL and GWAS suggest a single top SNP, that top SNP differs between eQTL and GWAS, and neither are MPRA hits (Fig. 5D and fig. S6A). Prior work has shown that a common splice variant in *ERAP2* results in nonsense-mediated decay (NMD) and allele-specific expression, which can cause an eQTL signal (*27*). However, the haplotype with this variant contains hundreds of other linked variants and harbors a second conditional *ERAP2* eQTL in Genotype-Tissue Expression (GTEx) LCLs. We evaluated eight of the 13 active variants from our MPRA by means of luciferase assay and found significant allelic differences at four of the eight loci (luciferase unpaired *t* test $P < 0.05$; rs1757538970, rs2549785, rs27298, and rs7713127) (fig. S7C). This suggests that *ERAP2* is regulated by a complex allelic structure that operates through gene expression and splicing.

Another colocalization was *PACSIN2*, which contained 13 variants and whose eQTL colocalized with platelet count. *PACSIN2* is an F-BAR domain protein involved in vascular and platelet homeostasis (*28*). We evaluated eight of the 13 allelic variants by means of luciferase assay and found significant effects at six of the eight loci (luciferase unpaired *t* test $P < 0.05$) (fig. S7D). Two of the variants (rs5751402 and rs9607970) were predicted to disrupt known TF binding motifs in directions consistent with their luciferase assay result (fig. S8B). The two TF were PAX5 and NFKB1, both of which are very highly expressed in LCLs, suggesting that rs5751402 and rs9607970 may function through disruption of NFKB1- and PAX5-binding sites.

Discussion

LD is a major barrier to identifying causal variants in genetic association studies. Furthermore, functional genomic annotations can be useful to prioritize likely causal variants, but many annotations are also inconclusive, unattainable, or unknown (*2*). In this study, we demonstrate that MPRAs provide a scalable platform with which to separate and map the regulatory activities of expression- and complex trait–associated natural genetic variants and highlight the limitations of existing approaches to variant interpretation and computational fine mapping. Across positional annotations and variant scores, we observed that both allelic MPRA hits and other tested variants were shifted relative to the corresponding genome-wide distributions. This demonstrates how functional predictions may readily distinguish eQTL regions from the genomic background while struggling to discriminate regulatory activity between highly linked allelic MPRA hits within the same region.

We found that multiple, tightly linked causal variants could be found under eQTL and GWAS loci. We identified that at least 17.7% of eQTL had more than one allelic hit. We further observed that most haplotype combinations exhibited additive effects, with 0.91% exhibiting nonadditivity. Using these data, we demonstrate the power of MPRA-based experimental fine mapping and report likely causal variants underlying hundreds of molecular and complex trait phenotypes, including a single variant underlying *ORMDL3*/asthma, three variants underlying *AHI1*/multiple sclerosis, and up to 13 variants each underlying *PACSIN2*/platelet count and *ERAP2*/Crohn's disease/inflammatory bowel disease.

REFERENCES AND NOTES

1. A. Buniello *et al.*, *Nucleic Acids Res.* **47** (D1), D1005–D1012 (2019).
2. V. Tam *et al.*, *Nat. Rev. Genet.* **20**, 467–484 (2019).
3. D. J. Schaid, W. Chen, N. B. Larson, *Nat. Rev. Genet.* **19**, 491–504 (2018).
4. A. Melnikov, X. Zhang, P. Rogov, L. Wang, T. S. Mikkelsen, *J. Vis. Exp.* **90**, e51719 (2014).
5. J. Ernst *et al.*, *Nat. Biotechnol.* **34**, 1180–1190 (2016).
6. M. Kircher *et al.*, *Nat. Commun.* **10**, 3583 (2019).
7. R. P. Patwardhan *et al.*, *Nat. Biotechnol.* **27**, 1173–1175 (2009).
8. R. Tewhey *et al.*, *Cell* **165**, 1519–1529 (2016).
9. J. C. Ulirsch *et al.*, *Cell* **165**, 1530–1545 (2016).
10. C. V. Weiss *et al.*, *eLife* **10**, e63713 (2021).
11. J. Choi *et al.*, *Nat. Commun.* **11**, 2718 (2020).
12. J. C. Klein *et al.*, *Nat. Commun.* **10**, 2434 (2019).
13. R. She, D. F. Jarosz, *Cell* **172**, 478–490.e15 (2018).
14. K. Renganaath *et al.*, *eLife* **9**, e62669 (2020).
15. Ž. Avsec *et al.*, *Nat. Methods* **18**, 1196–1203 (2021).
16. X. Li *et al.*, *Nat. Genet.* **52**, 969–983 (2020).
17. T. Lappalainen *et al.*, *Nature* **501**, 506–511 (2013).
18. GTEx Consortium, *Science* **369**, 1318–1330 (2020).
19. J. C. Kwasnieski, I. Mogno, C. A. Myers, J. C. Corbo, B. A. Cohen, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 19498–19503 (2012).
20. J. E. Powell *et al.*, *PLOS Genet.* **9**, e1003502 (2013).
21. V. Hivert *et al.*, *Am. J. Hum. Genet.* **108**, 786–798 (2021).
22. A. N. Barbeira *et al.*, *Genome Biol.* **22**, 49 (2021).
23. D. J. Verlaan *et al.*, *Am. J. Hum. Genet.* **85**, 377–393 (2009).
24. A. Rathod *et al.*, *Epigenet. Insights* **13**, 2516865720923395 (2020).
25. B. J. Kaskow *et al.*, *Neurol. Neuroimmunol. Neuroinflamm.* **5**, e414 (2017).
26. K. Christodoulou *et al.*, *Gut* **62**, 977–984 (2013).
27. A. M. Andrés *et al.*, *PLOS Genet.* **6**, e1001157 (2010).
28. A. J. Begonja *et al.*, *Blood* **126**, 80–88 (2015).
29. N. Abell, nsabell/mpra-v2: FineMapMPRA. *Zenodo* (2022); doi: 10.5281/zenodo.5921041.

ACKNOWLEDGMENTS

We thank members of the S. Montgomery laboratory for general guidance and feedback on this work and members of the M. Bassik laboratory for experimental advice. We also thank R. Tewhey and M. Love for experimental and statistical modeling advice, respectively, and N. Cyr for assistance with figures. **Funding:** N.S.A. is supported by the Stanford Department of Genetics T32 training grant and the Joint Institute for Metrology in Biology (JIMB) training program. E.G. is funded by the National Science Foundation Graduate Research Fellowship Program grant DGE-1656518. S.B.M. is supported by National Institutes of Health grants R01AG066490, R01MH125244, U01HG009431 (ENCODE), R01HL142015 (TOPMed), and R01HG008150 (NoVa). This work in part used supercomputing resources provided by the Stanford Genetics Bioinformatics Service Center, supported by National Institutes of Health S10 Instrumentation Grant S10OD023452. **Author contributions:** N.S.A. and S.B.M. conceived and designed the study. N.S.A., M.K.D., E.G., and K.S.S. performed all experiments, including the MPRA and luciferase assays. N.S.A., M.K.D., and M.J.G. designed oligonucleotide libraries. N.S.A. and Z.H. conducted statistical and bioinformatic analyses of sequencing data. N.S.A. and S.B.M wrote the manuscript, with contributions from all authors. **Competing interests:** S.B.M. has consulting agreements with MyOme, Biomarin, and Tenaya Therapeutics. All other authors report no competing interests. **Data and materials availability:** Sequencing data are available through the Gene Expression Omnibus under accession no. GSE174534. All code and supplementary tables are available through Zenodo (*29*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj5117
Materials and Methods
Figs. S1 to S9
Tables S1 to S9
References (*30–47*)
MDAR Reproducibility Checklist

17 May 2021; resubmitted 19 October 2021
Accepted 17 February 2022
10.1126/science.abj5117

CELL DEATH

Copper induces cell death by targeting lipoylated TCA cycle proteins

Peter Tsvetkov[1]*, Shannon Coy[2,3,4,5], Boryana Petrova[5,6], Margaret Dreishpoon[1], Ana Verma[2,3,4,5], Mai Abdusamad[1], Jordan Rossen[1], Lena Joesch-Cohen[1], Ranad Humeidi[1], Ryan D. Spangler[1], John K. Eaton[1], Evgeni Frenkel[7], Mustafa Kocak[1], Steven M. Corsello[1,5,8], Svetlana Lutsenko[9], Naama Kanarek[1,5,6], Sandro Santagata[2,3,4,5,10], Todd R. Golub[1,5,11,12]*

Copper is an essential cofactor for all organisms, and yet it becomes toxic if concentrations exceed a threshold maintained by evolutionarily conserved homeostatic mechanisms. How excess copper induces cell death, however, is unknown. Here, we show in human cells that copper-dependent, regulated cell death is distinct from known death mechanisms and is dependent on mitochondrial respiration. We show that copper-dependent death occurs by means of direct binding of copper to lipoylated components of the tricarboxylic acid (TCA) cycle. This results in lipoylated protein aggregation and subsequent iron-sulfur cluster protein loss, which leads to proteotoxic stress and ultimately cell death. These findings may explain the need for ancient copper homeostatic mechanisms.

The requirement of copper as a cofactor for essential enzymes has been recognized across the animal kingdom, spanning bacteria to human cells (*1*). However, intracellular copper concentrations are kept at extraordinarily low levels by active homeostatic mechanisms that work across concentration gradients to prevent the accumulation of free intracellular copper that is detrimental to cells (*1–4*). Whereas the mechanisms of toxicity of other essential metals, such as iron, are well established, the mechanisms of copper-induced cytotoxicity remain unclear (*5–7*).

Copper ionophores are copper-binding small molecules that shuttle copper into the cell and are thereby useful tools to study copper toxicity (*8*, *9*). Multiple lines of evidence indicate that the mechanism of copper ionophore–induced cell death involves intracellular copper accumulation and not the effect of the small molecule chaperones themselves. Multiple, structurally distinct small molecules that bind copper share killing profiles across hundreds of cell lines [Fig. 1A, fig. S1A, and (*5*, *10*)]. Structure-function relationship experiments show that modifications that abrogate the copper binding capacity of these compounds result in loss of cell killing (*5*), and copper chelation eliminates the cytotoxicity of the compounds [fig. S1, B to C, and (*5*)].

A clear picture of the mechanisms underlying copper-induced toxicity has not yet emerged, with contradictory reports suggesting either the induction of apoptosis (*11*, *12*), caspase-independent cell death (*5*, *7*, *13*), reactive oxygen species (ROS) induction (*14–16*), or inhibition of the ubiquitin-proteasome system (*17–19*). The cross-kingdom efficacy of copper-binding molecules as cell death inducers suggests that they target evolutionarily conserved cellular machinery, but such mechanisms have yet to be elucidated.

To further establish whether copper ionophore cytotoxicity is dependent on copper itself, we analyzed the killing potential of the potent copper ionophore elesclomol. The source of copper in cell culture medium is serum (fig. S1Q), and accordingly, cells grown in the absence of serum were resistant to elesclomol. By contrast, elesclomol sensitivity was completely restored by the addition of copper in a 1:1 ratio (fig. S1, E and F). Copper supplementation similarly sensitized cells to treatment with six structurally distinct copper ionophores (Fig. 1B and fig. S1, G to P), but supplementation with other metals, including iron, cobalt, zinc, and nickel, failed to potentiate cell death (Fig. 1B and fig. S1, G to P). Consistent with this observation, depletion of the endogenous intracellular copper chelator glutathione, using buthionine sulfoximine (BSO), sensitized cells to elesclomol-copper–induced cell death (fig. S1Q), whereas chelation of copper with tetrathiomolybdate (TTM) rescued killing (fig. S1, B and C) and chelators of other metals had no effect (fig. S1R). Lastly, a 2-hour pulse treatment with potent copper ionophores (elesclomol, disulfiram, and NSC319726) resulted in a ~5- to 10-fold increase in levels of intracellular copper but not zinc (fig. S2A). These results suggest that copper ionophore–induced cell death is primarily dependent on intracellular copper accumulation.

Copper ionophores induce a distinct form of regulated cell death

We first asked whether copper ionophore–mediated cell death is regulated and specifically whether short-term exposure leads to irrevocable, subsequent cell cytotoxicity. Pulse treatment with the copper ionophore elesclomol at concentrations as low as 40 nM for only 2 hours resulted in a 15- to 60-fold increase in intracellular copper levels (fig. S2, B and C) that triggered cell death more than 24 hours later (Fig. 1C). This result suggests that copper-mediated cell death is indeed regulated.

Cell death involves signaling cascades and molecularly defined effector mechanisms (*20*) that involve proteins and lipids such as those characteristic of apoptosis (*21*), necroptosis (*22*), pyroptosis (*23*), and ferroptosis (*24*), which is a recently discovered iron-dependent cell death pathway. Previous reports suggested that elesclomol induces ROS-dependent apoptotic cell death (*6*, *12*), but elesclomol-induced cell death did not involve either the cleavage or activation of caspase 3 activity, the hallmark of apoptosis (*25*) (Fig. 1, D and E, and fig. S2D). Similarly, elesclomol killing potential was maintained when the key effectors of apoptosis, BAX and BAK1, were knocked out (Fig. 1F and fig. S2, E to I) or when cells were cotreated with pan-caspase inhibitors (Z-VAD-FMK and Boc-D-FMK)

---

[1]Broad Institute of Harvard and MIT, Cambridge, MA, USA. [2]Laboratory of Systems Pharmacology, Department of Systems Biology, Boston, MA, USA. [3]Ludwig Center at Harvard, Harvard Medical School, Boston, MA, USA. [4]Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA. [5]Harvard Medical School, Boston, MA, USA. [6]Department of Pathology, Boston Children's Hospital, Boston, MA USA. [7]Whitehead Institute and Massachusetts Institute of Technology, Cambridge, MA, USA. [8]Department of Medical Oncology, Dana Farber Cancer Institute, Boston, MA, USA. [9]Department of Physiology, Johns Hopkins Medical Institutes, Baltimore, MD, USA. [10]Department of Pathology, Dana Farber Cancer Institute, Boston, MA, USA. [11]Department of Pediatric Oncology, Dana Farber Cancer Institute, Boston, MA, USA. [12]Division of Pediatric Hematology/Oncology, Boston Children's Hospital, Boston, MA, USA.
***Corresponding author. Email: ptsvetko@broadinstitute.org (P.T.); golub@broadinstitute.org (T.R.G.)**

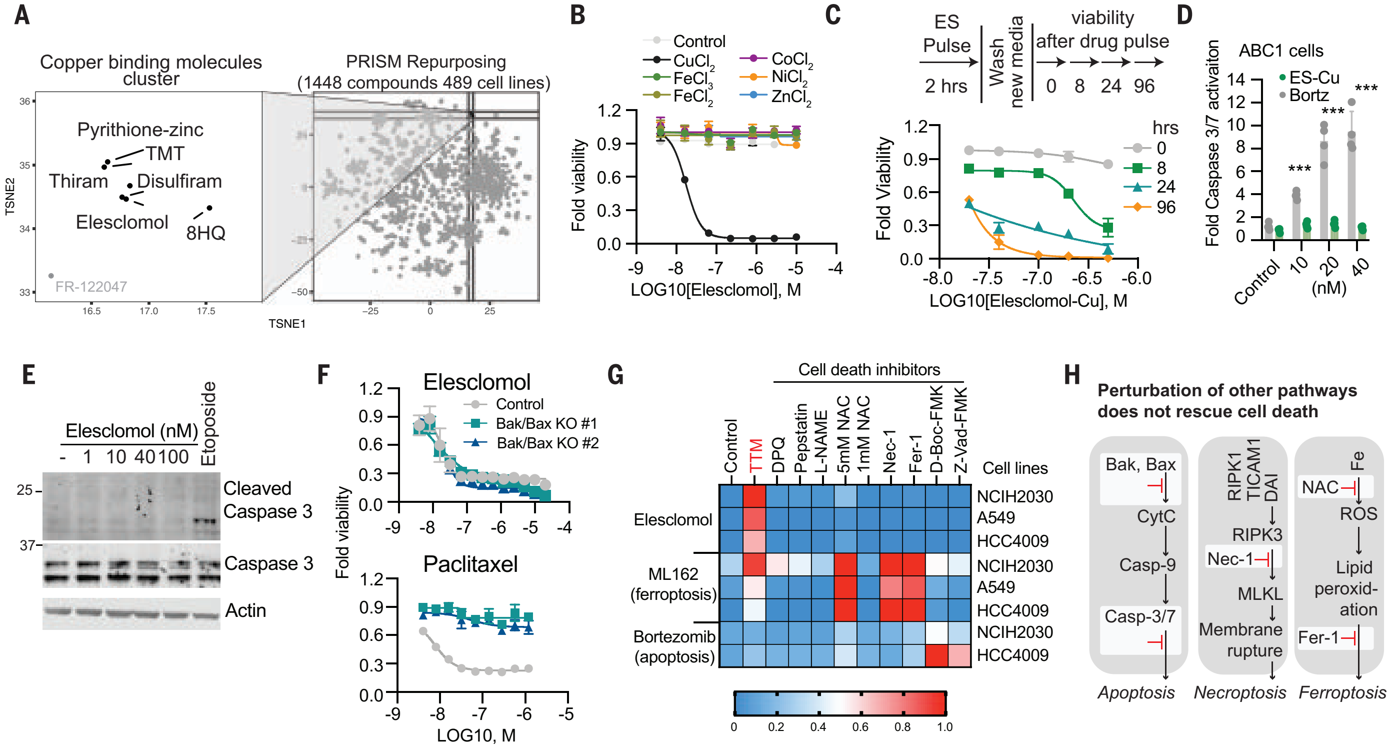


Fig. 1. Copper ionophore–induced cell death is nonapoptotic, nonferroptotic, and non-necroptotic. (**A**) PRISM (profiling relative inhibition simultaneously in mixtures) Repurposing secondary screen: Growth-inhibition estimates for 1448 drugs against 489 cell lines. TMT, tetramethylthiuram monosulfide; TSNE, t-distributed stochastic neighbor embedding. (**B**) Viability of cells (MON) after treatment with elesclomol with or without 10 μM of indicated metals. (**C**) Viability of ABC1 cells was assessed at the indicated times after elesclomol-Cu (1:1 ratio) pulse treatment and growth in fresh media. ES, elesclomol. (**D**) Caspase 3/7 cleavage in ABC1 cells 16 hours after indicted treatments (fold change over control). (**E**) Western blot analysis of G402 cells treated with the indicated concentrations of elesclomol or 25 μM etoposide for 6 hours. (**F**) Viability of HMC18 cells or two HMC18 clones with Bax/Bak deleted after treatment with elesclomol-CuCl$_2$ (1:1) (top) and paclitaxel (bottom). (**G**) Heatmap of viability of cells pretreated overnight with 20 μM necrostatin-1, 10 μM ferrostatin-1, 1 mM *N*-acetylcysteine (NAC), 5 mM *N*-acetylcysteine, 30 μM Z-VAD-FMK, 50 μM D-Boc-FMK, 20 μM TTM, 300 μM L-NAME, 1 μM pepstatin A, or 10 μM DPQ and then treated with either 30 nM elesclomol-CuCl$_2$ (1:1), 1 μM ML162 (GPX4 inhibitor), or 40 nM bortezomib for 72 hours (average of three replicates). (**H**) Schematic diagram of apoptosis, necroptosis, and ferroptosis. Inhibited pathways are marked in red. For (B), (D), and (F), data are means ± SD, with $n \geq 3$. For (D) and (E), media were supplemented with 1 μM CuCl$_2$.

(Fig. 1G), again indicating that the copper-induced cell death is distinct from apoptosis. Furthermore, treatment with inhibitors of other known cell death mechanisms—including ferroptosis (ferrostatin-1), necroptosis (necrostatin-1), and oxidative stress (*N*-acetyl cysteine)—all failed to abrogate copper ionophore–induced cell death (Fig. 1G), suggesting a mechanism distinct from known cell death pathways (Fig. 1H).

Mitochondrial respiration regulates copper ionophore–induced cell death

One hint to the pathways that mediate copper ionophore–induced cell death is the observation that cells that are more reliant on mitochondrial respiration are nearly 1000-fold more sensitive to copper ionophores than cells undergoing glycolysis (Fig. 2A and fig. S3, A to F). Treatment with mitochondrial antioxidants, fatty acids, and inhibitors of mitochondrial function had a very distinct effect on the sensitivity to copper ionophores as compared with sensitivity to the ferroptosis-inducing GPX4 inhibitor ML162 (Fig. 2B). Furthermore, inhibitors of complexes I and II of the electron transport chain (ETC) as well as inhibitors of mitochondrial pyruvate uptake attenuated cell death with no effect on ferroptosis (Fig. 2B). Importantly, the mitochondrial uncoupler FCCP had no effect on copper toxicity, suggesting that mitochondrial respiration, not adenosine triphosphate (ATP) production, is required for copper-induced cell death (Fig. 2C). Consistent with this finding, growing cells in hypoxic conditions (1% O$_2$) attenuated copper ionophore–induced cell death, whereas forced stabilization of the hypoxia-inducible factor (HIF) pathway with the HIF prolyl hydroxylase inhibitor FG-4592 under normoxic conditions (21% O$_2$) did not (Fig. 2D and fig. S3, G to J), further emphasizing the role of cellular respiration in mediating copper-induced cell death. However, treatment with copper ionophores did not induce significant reduction in basal or ATP-linked respiration but did significantly reduce the spare capacity of respiration (Fig. 2E and fig. S3, K to N), suggesting that copper does not target the ETC directly but rather components of the TCA cycle. In support of this, metabolite profiling of cells pulse-treated with elesclomol showed a time-dependent increase in metabolite dysregulation of many TCA cycle–associated metabolites in elesclomol-sensitive ABC1 cells but not in elesclomol-resistant A549 cells (fig. S3, O to S, and table S1). These results establish a link between copper ionophore–induced cell death and mitochondrial metabolism, not the ETC (Fig. 2F), leading us to further elucidate the precise connection between copper and the TCA cycle.

FDX1 and protein lipoylation are the key regulators of copper ionophore–induced cell death

To identify the specific metabolic pathways that mediate copper toxicity, we performed genome-wide CRISPR-Cas9 loss-of-function screens to identify the genes involved in copper ionophore–induced death. To maximize the generalizability of the screen, we focused on the intersection of two structurally distinct copper-loaded ionophores (elesclomol and the active form of disulfiram, diethyldithiocarbamate) (Fig. 3, A to C, and table S2). Killing by both compounds was rescued by knockout

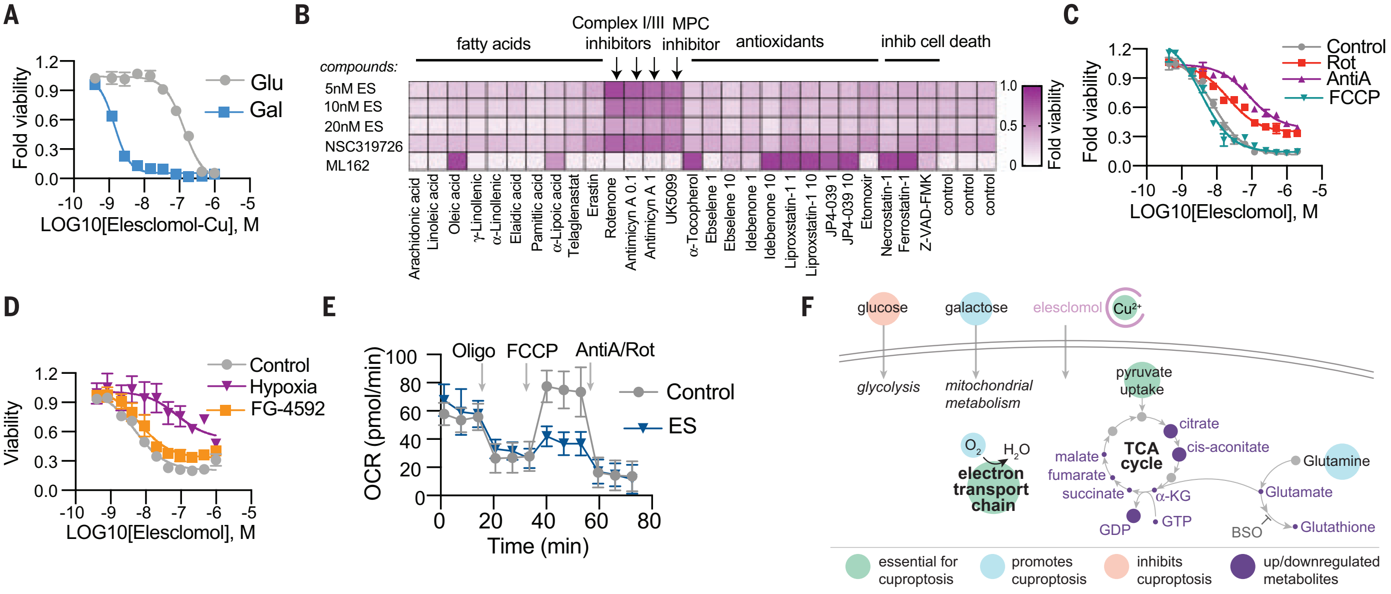


Fig. 2. Mitochondria respiration regulates copper ionophore–induced cell death. (**A**) Viability of NCIH2030 cells grown in media containing either glucose or galactose treated with elesclomol-Cu (ratio 1:1). (**B**) Viability of ABC1 cells pretreated with the indicated compounds (*x* axis) and then treated with elesclomol, NSC319726, or ML162 (*y* axis). The average of at least three replicates is plotted and color coded. MPC, mitochondrial pyruvate carrier. (**C**) Viability of ABC1 cells pretreated with 0.1 μM rotenone (Rot), 0.1 μM antimycin A (AntiA), or 1 μM FCCP and then treated with elesclomol. (**D**) Viability of ABC1 cells grown in control (21% O_2), hypoxia (1% O_2), or 50 μM FG-4592 (21% O_2) after treatment with elesclomol. (**E**) The oxygen consumption rate (OCR) was detected after treatment with 2.5 nM elesclomol (with 1 μM $CuCl_2$ in media) for 16 hours of ABC1 cells before (basal) and after the addition of oligomycin (ATP-linked), the uncoupler FCCP (maximal), or the electron transport inhibitor antimycin A/rotenone (baseline) (mean ± SD, $n = 8$). (**F**) Schematic of metabolites altered after elesclomol treatment of ABC1 cells. (Purple circles mark metabolites that change in abundance, as detailed in table S1. Larger circles are metabolites that are up-regulated, and smaller circles are metabolites that are down-regulated). GDP, guanosine diphosphate; GTP, guanosine triphosphate. For (A), (C), and (D), data are means ± SD, with $n \geq 3$.

of seven genes (Fig. 3A, marked in blue), including *FDX1* [which encodes a reductase known to reduce Cu^{2+} to its more toxic form, Cu^{1+}, and to be a direct target of elesclomol (*5*)] and six genes that encode either components of the lipoic acid pathway [lipolytransferase 1 (LIPT1), lipoyl synthase (LIAS), and dihydrolipoamide dehydrogenase (DLD)] or protein targets of lipoylation [the pyruvate dehydrogenase (PDH) complex, including dihydrolipoamide *S*-acetyltransferase (DLAT), pyruvate dehydrogenase E1 subunit alpha 1 (PDHA1), and pyruvate dehydrogenase E1 subunit beta (PDHB)] (*26*) (Fig. 3D). These observations were validated by an independent knockout screen that focused on 3000 metabolic enzymes (*27*) (Fig. 3E; fig. S4, A and B; and table S2). In addition, the metabolic enzyme screen showed that genetic suppression of complex I also rescued cells from copper-induced death (Fig. 3E; fig. S4, A and B; and table S2), consistent with our finding that chemical inhibitors of the ETC block cell death mediated by copper ionophores (Fig. 2).

Individual gene knockout studies further confirmed that deletion of *FDX1* and *LIAS* conferred resistance to copper-induced cell death (Fig. 3, F and G, and fig. S4, C to K), further strengthening a functional link between FDX1, the protein lipoylation machinery, and copper toxicity. Moreover, FDX1 deletion resulted in consistent resistance to a number of copper ionophores (disulfiram, NSC319726, thiram, 8-HQ, and Zn-pyrithione) (fig. S5, A to G) but showed no substantial effects on either caspase activation (fig. S5, H to K), the potency of apoptosis-inducers (bortezomib, paclitaxel, and topotecan), or the ferroptosis-inducer ML162 (fig. S5, L to O). Of note, the strong connection between copper-mediated cell death and both FDX1 expression and protein lipoylation was lost at high concentrations of elesclomol (>40 nM) (fig. S4, L to P), thereby suggesting that off-target mechanisms of cell death may occur at such concentrations and possibly explaining conflicting mechanisms of action reported in the literature (*5*, *7*, *14*–*16*, *19*, *28*, *29*). Moreover, whereas the most copper-selective compounds (e.g., elesclomol, disulfiram, and NSC319726) lost killing activity when cells were grown under glycolytic conditions, compounds with more promiscuous metal-binding compounds (e.g., pyrithione and 8-HQ) killed independent of metabolic state. This result is consistent with copper's distinctive connection to mitochondrial metabolism-mediated protein lipoylation. (Fig. 2A and figs. S3, A to F, and S5, A to G).

FDX1 is an upstream regulator of protein lipoylation

Protein lipoylation is a highly conserved lysine posttranslational modification that is known to occur on only four enzymes, all of which involve metabolic complexes that regulate carbon entry points to the TCA cycle (*26*, *30*). These include dihydrolipoamide branched chain transacylase E2 (DBT), glycine cleavage system protein H (GCSH), dihydrolipoamide *S*-succinyltransferase (DLST), and DLAT, an essential component of the PDH complex. Lipoylation of these proteins is known to be required for enzymatic function (*30*) (Fig. 3D). Our findings that knockout of either FDX1 or lipoylation-related enzymes rescues cells from copper toxicity led us to explore whether FDX1 might be an upstream regulator of protein lipoylation. To test this hypothesis, we performed three analyses. First, we looked for evidence of coordinated dependencies across the Cancer Dependency Map (www.depmap.org), a resource of genome-wide CRISPR-Cas9 knockout screens in hundreds of cancer cell lines. Genes showing similar patterns of viability effects, even if subtle, suggest that they have shared function or regulation. Notably, FDX1 and components of the lipoic acid pathway were highly correlated in their viability effects across the cell line panel ($p < 0.0001$; Fig. 4A).

Second, we performed immunohistochemical staining for FDX1 and lipoic acid in 208 human tumor specimens and semiquantitative light-microscopic scoring by two independent pathologists. Expression of FDX1 and lipoylated

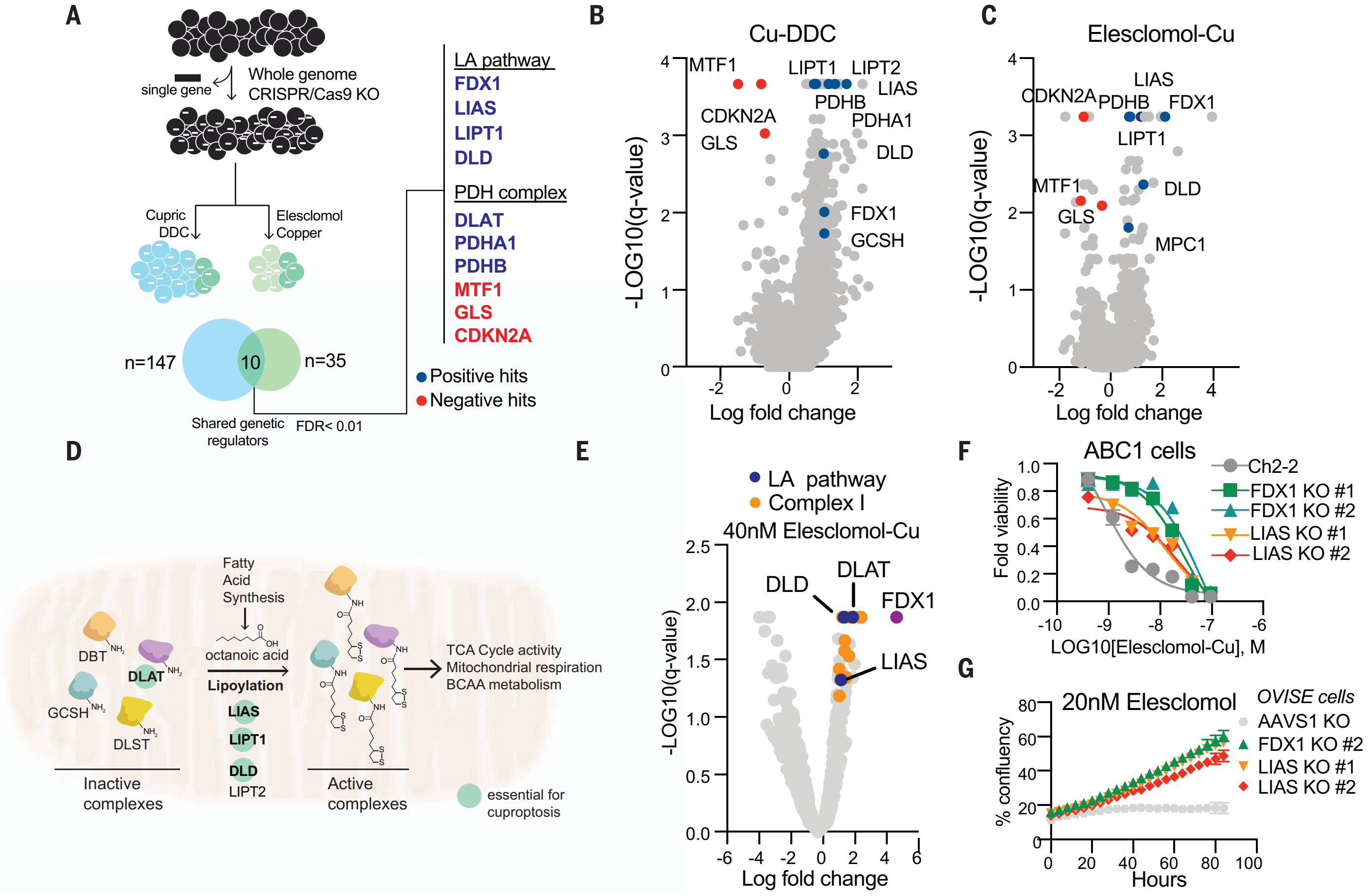


Fig. 3. FDX1 and lipoic acid genes are critical mediators of copper ionophore–induced cell death. (**A**) Whole-genome CRIPSR-Cas9 positive selection screen using two copper ionophores (Cu-DDC and elesclomol-copper) in OVISE cells (schematic on the left). Overlapping hits with a false discovery rate (FDR) score <0.01 were analyzed (right). Positive hits (resistance) are marked in blue, and negative hits (sensitizers) are marked in red. (**B** and **C**) Summary scatter of the results of the screen in (A) for Cu-DDC (B) or elesclomol-Cu (C). (**D**) Schematic of the lipoic acid pathway. Genes that scored in our genetic screens are marked as essential for copper-induced cell death. BCAA, branched-chain amino acid. (**E**) Summary scatter plot indicating top hits in the metabolism gene–focused CRISPR-Cas9 gene knockout screen of A549 cells treated with 40 nM of elesclomol-Cu(II) (1:1 ratio). Genes associated with the lipoic acid pathway are marked in blue, complex I–related genes in orange, and FDX1 in purple. (**F**) Viability of ABC1 cells with CRIPSR-Cas9 deletion of *LIAS* or *FDX1* after treatment with elesclomol in the presence of 1 μM $CuCl_2$ in the media. (**G**) Growth curve measurements of OVISE cells with CRIPSR-Cas9 deletion of *LIAS* and *FDX1* in the presence of 20 nM elesclomol. For (F) and (G), data are means ± SD, with $n \geq 3$.

proteins was highly correlated ($p < 0.0001$; Fig. 4, B and C, and fig. S6, A to D). Third, we determined whether FDX1 knockout affected protein lipoylation using a lipoic acid–specific antibody as a measure of DLAT and DLST lipoylation. FDX1 knockout resulted in complete loss of protein lipoylation, as measured by either immunoblot or immunohistochemistry (Fig. 4D and fig. S6, E to I), and also led to a significant drop in cellular respiration similar to the levels observed with the deletion of LIAS itself (Fig. 4E and fig. S6J). Furthermore, metabolite profiling after deletion of FDX1 led to an accumulation of pyruvate and α-ketoglutarate and depletion of succinate, as would be expected when protein lipoylation is compromised because of inhibition of the TCA cycle at PDH and α-ketoglutarate dehydrogenase (*31*) (Fig. 4F and table S3). Also of interest, we observed an increase in *S*-adenosylmethionine (SAM), a key substrate of LIAS in the lipoic acid pathway, consistent with FDX1 being a previously unrecognized upstream regulator of protein lipoylation.

Copper directly binds and induces the oligomerization of lipoylated DLAT

The experiments described above establish a connection between copper toxicity and protein lipoylation but do not establish a direct mechanistic link. We hypothesized that copper might directly bind to lipoylated proteins, a possibility suggested by the observation that copper binds free lipoic acid with a measured dissociation constant of 10^{-17} (*32*). To test this hypothesis, we purified DLAT and DLST from cell lysates and found that these proteins bound to copper-charged resin but not to cobalt or nickel resins (Fig. 5A). When protein lipoylation was abrogated by FDX1 deletion (Fig. 4), DLAT and DLST no longer bound copper (Fig. 5B), suggesting that the lipoyl moiety is required for copper binding.

We noted that copper binding to lipoylated proteins did not simply lead to loss of function, given that deletion of the proteins rescues (not phenocopies) copper ionophore treatment. We proposed that copper binding to lipoylated TCA cycle proteins results in a toxic gain of function. Interestingly, we found that the copper binding to lipoylated TCA cycle proteins resulted in lipoylation-dependent oligomerization of DLAT that was detectable by nondenaturing gel electrophoresis (Fig. 5, B and C). Similarly, treatment of elesclomol-sensitive cells increased levels of DLAT oligomers and insoluble DLAT, whereas treatment of elesclomol-insensitive cell lines or FDX1 knockout (FDX1 KO) cells resulted in DLAT oligomerization only at much higher concentrations (Fig. 5, D and E, and fig. S7, A and B). Treatment with the strong

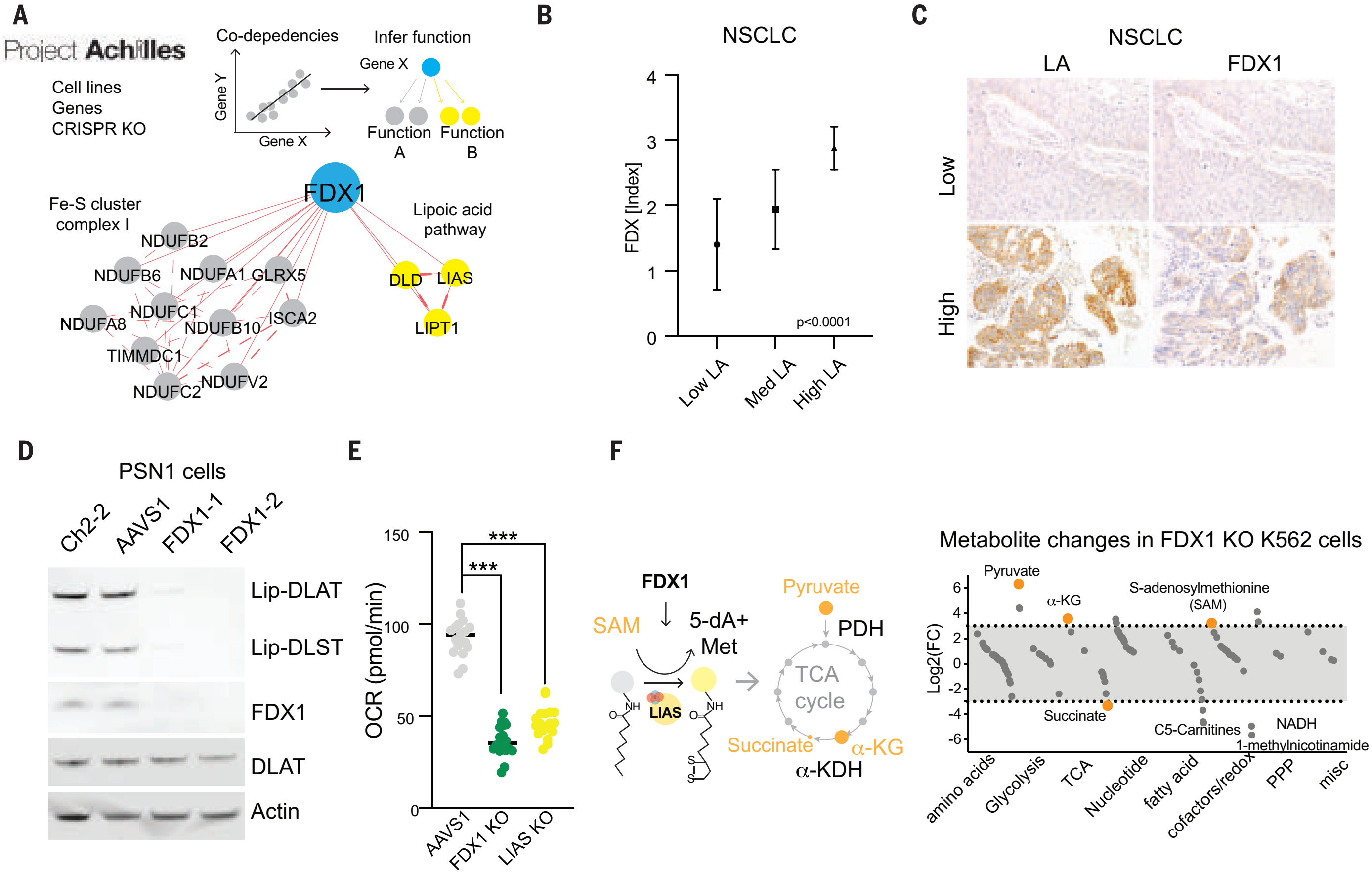


Fig. 4. FDX1 is an upstream regulator of protein lipoylation. (**A**) Correlation analysis of gene dependencies taken from the Achilles project. Presented is the gene network that correlates with *FDX1* deletion using FIREWORKS (*51*). The correlating genes are marked by their described functionality (lipoic acid pathway, and mitochondria complex I and Fe-S cluster regulation). (**B**) A tissue microarray of non–small cell lung carcinoma (NSCLC) ($n = 57$) was stained with lipoic acid (LA), and FDX immunohistochemistry (IHC) and expression was scored semiquantitatively by two pathologists (S.C. and S.S.), showing a strong direct correlation between LA and FDX expression (mean ± SD; $p < 0.0001$). (**C**) Representative cases of NSCLC with correlated low (top row) and high (bottom row) expression of LA and FDX1 by IHC. (**D**) Immunoblot of lipoylated proteins, FDX1, DLAT, and actin from extracts of PSN1 cells with deletion of *FDX1*. (**E**) Basal OCR as measured in PSN1 cells with CRIPSR-Cas9 deletion of *FDX1*, *LIAS*, or *AAVS1* genes (unpaired *t* test, ***$p < 0.001$). (**F**) Plot of the average log2 fold change (FC) in metabolites between *FDX1* KO and *AAVS1* control K562 cell lines separated by functional annotations. Metabolites relevant to the lipoic acid pathway are marked in orange. α-KDH, α-ketoglutarate dehydrogenase; α-KG, α-ketoglutarate; NADH, reduced nicotinamide adenine dinucleotide; PPP, pentose phosphate pathway.

reducing agent tris(2-carboxyethyl)phosphine (TCEP) and boiling eliminated the oligomeric form of DLAT (fig. S7, C to E), suggesting that the aggregates are disulfide-bond dependent. We confirmed these findings by immunofluorescence, observing pronounced induction of DLAT foci by short-pulse elesclomol treatment, whereas such foci were diminished in FDX1 KO, lipoylation-deficient cells (Fig. 5, F to H, and fig. S7F). These findings support a model where the toxic gain of function of lipoylated proteins after exposure to copper ionophores is mediated at least in part by their aberrant oligomerization. Mass spectrometric analysis also revealed that copper ionophore treatment leads to loss of Fe-S cluster proteins (Fig. 6, A and B) in an FDX1-dependent manner (Fig. 6C) as well as the induction of proteotoxic stress (Fig. 6, B and C; fig. S8A; and table S4). These findings are consistent with the observation in bacteria and yeast that copper can destabilize Fe-S–containing proteins (*33*–*35*). Whether such loss of Fe-S cluster proteins contributes to the copper ionophore death phenotype remains to be determined.

Copper-induced death mechanisms are shared by genetic models of copper homeostasis dysregulation

The experiments described to this point used copper ionophores to overcome the homeostatic mechanisms that normally keep intracellular copper concentration low. These mechanisms include the copper importer SLC31A1 (CTR1) and the copper exporters ATP7A and ATP7B, which are encoded by genes that are mutated in the copper dysregulation syndromes Menke's disease and Wilson's disease, respectively (*3*, *36*) (Fig. 6D). To explore whether the mechanisms of copper toxicity associated with copper ionophore treatment are shared by these naturally occurring disorders of copper homeostasis, we examined three experimental models. First, we overexpressed SLC31A1 in human embryonic kidney (HEK) 293T and ABC1 cells, which dramatically increased sensitivity to physiological concentrations of copper (Fig. 6E and fig. S8B). Consistent with our findings using copper ionophores in wild-type cells, copper supplementation resulted in an overall reduction in proteins involved in mitochondrial respiration (fig. S8C and table S4), reduced protein lipoylation, reduced levels of Fe-S cluster proteins, and increased levels of HSP70 (Fig. 6F). Importantly, ferroptosis, necroptosis, and apoptosis inhibitors did not prevent copper-induced cell death in cells overexpressing SLC31A1 (Fig. 6G and fig. S8, D to F), whereas copper chelators, FDX1 KO, and LIAS KO each partially rescued cells from copper-induced cell death (Fig. 6, G and H, and fig. S8, D to G). In addition, depletion of the natural intracellular copper chaperone glutathione resulted in copper-dependent cell death (Fig. 6I), which was associated with reduced lipoylation

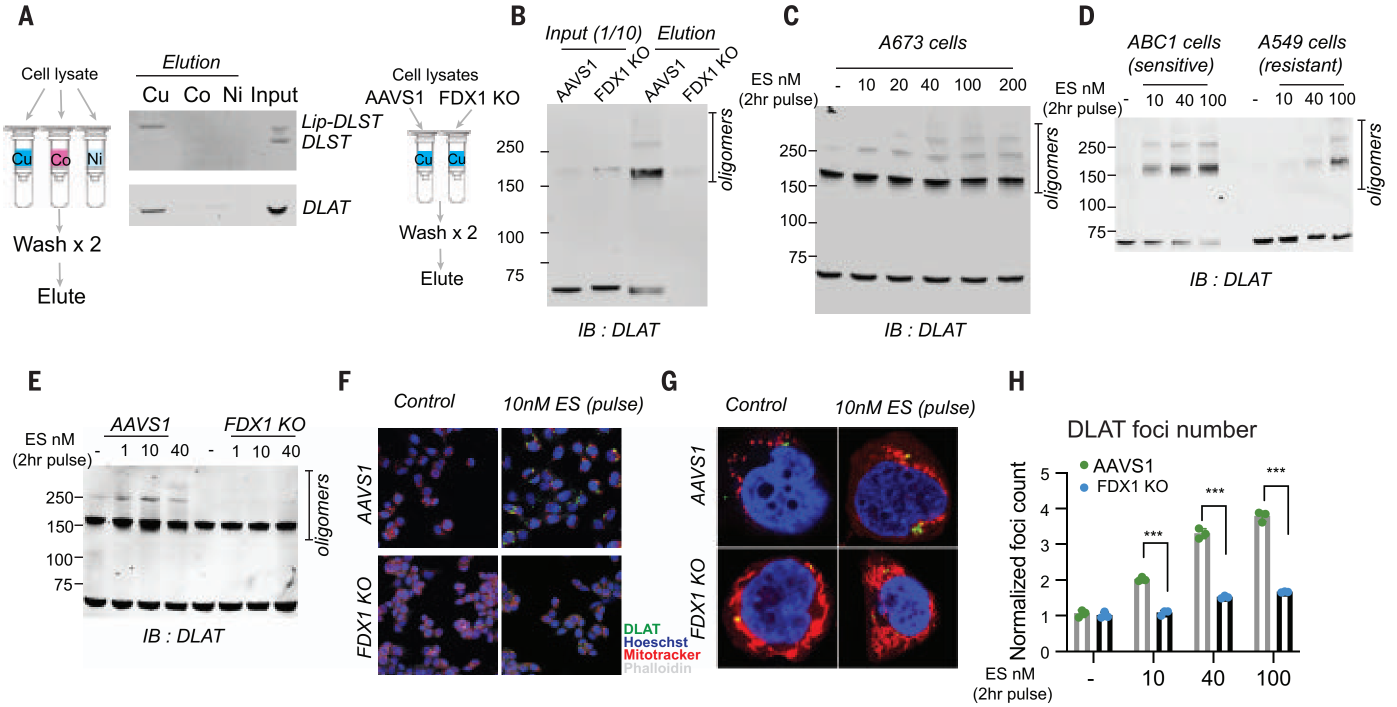


Fig. 5. Copper directly binds and promotes the oligomerization of lipoylated DLAT. (**A**) The binding of the indicated proteins to copper (Cu), cobalt (Co), and nickel (Ni) was assessed by immunoblot analysis of eluted proteins from the indicated metal-loaded resins. (**B**) Copper binding was assessed by loading cell lysates from either ABC1 AAVS1 or FDX1 KO cells on copper-loaded resin followed by washing and analysis of the eluted proteins. Input and eluted proteins are presented. IB, immunoblot. (**C**) Protein content was analyzed in A673 cells that were pulse-treated (2 hours) with the indicated concentrations of elesclomol. (**D**) Protein content was analyzed in ABC1 and A549 cells that were pulse-treated with the indicated concentrations of elesclomol. (**E** to **H**) AAVS1 control or FDX1 KO ABC1 cells were pulse-treated with the indicated concentrations of elesclomol for 2 hours; protein oligomerization was analyzed after 24 hours by immunoblotting (E) and both wide-field (F) and confocal (G) immunofluorescence imaging (green, DLAT–; red, Mitotracker; blue, Hoechst; white, phalloidin). (H) Foci were segmented and quantified in each condition [n = 3, and multiple unpaired t test analysis was conducted with the desired FDR(Q) = 1%; ***q < 0.001]. For (C) to (H), media were supplemented with 1 μM $CuCl_2$.

and increased DLAT oligomerization (fig. S8, I and J) that was attenuated by FDX1 and LIAS KO, just as was observed with copper ionophore treatment of SCL31A1 wild-type cells (Fig. 6J and fig. S8H).

Lastly, we used a mouse model of Wilson's disease in which Atp7b deletion leads to intracellular copper accumulation and cell death with increasing animal age (*37*, *38*). Comparing the livers of aged Atp7b-deficient (*Atp7b*$^{-/-}$) mice with those of Atp7b heterozygous (*Atp7b*$^{-/+}$) and wild-type control mice, we observed loss of lipoylated and Fe-S cluster proteins, as well as an increase in Hsp70 abundance (Fig. 6K and table S5). These findings in mouse models of copper toxicity suggest that copper overload results in the same cellular effects as those induced by copper ionophores. Taken together, our data support a model whereby excess copper promotes the aggregation of lipoylated proteins and destabilization of Fe-S cluster proteins that results in proteotoxic stress and ultimately cell death (Fig. 6L).

Discussion

Copper is a double-edged sword: It is essential as a cofactor for enzymes across the animal kingdom, and yet even modest intracellular concentrations can be toxic, resulting in cell death (*2*). Genetic variation in copper homeostasis results in life-threatening disease (*39*, *40*), and both copper ionophores (*11*, *17*, *41*, *42*) and copper chelators (*43*–*46*) have been suggested as anticancer agents. However, the mechanism by which copper overload leads to cell death has been obscure. Here, we have shown that copper toxicity occurs by a mechanism distinct from all other known mechanisms of regulated cell death, including apoptosis, ferroptosis, pyroptosis, and necroptosis. We therefore propose that this previously uncharacterized cell death mechanism be termed cuproptosis.

We have shown that copper-induced cell death is mediated by an ancient mechanism: protein lipoylation. Notably few mammalian proteins are known to be lipoylated, and these are concentrated in the TCA cycle, where lipoylation is required for enzymatic function (*26*, *47*). Our study explains the relationship between mitochondrial metabolism and the sensitivity to copper-mediated cell death: Respiring, TCA-cycle active cells have increased levels of lipoylated TCA enzymes (in particular, the PDH complex), and the lipoyl moiety serves as a direct copper binder, resulting in lipoylated protein aggregation, loss of Fe-S cluster–containing proteins, and induction of HSP70, reflective of acute proteotoxic stress. The targets of copper-induced toxicity that we describe here in human cancer cells (lipoylation and Fe-S cluster proteins) are evolutionarily conserved from bacteria to humans, suggesting that copper-induced cell death might be also used by microorganisms where copper ionophores are naturally synthesized and exhibit antimicrobial activity (*48*, *49*).

In the case of genetic disorders of copper homeostasis (Wilson's disease and Menke's disease), copper chelation is an effective form of therapy (*50*). In cancer, however, the exploitation of copper toxicity has been less successful (*42*). Copper ionophores, including elesclomol, have been tested in clinical trials, but such testing occurred without the benefit of either a biomarker of the appropriate patient population or an understanding of the drug's mechanism of action. A phase 3 combination clinical trial of elesclomol in patients with melanoma showed a lack of efficacy in this unselected population, but a post hoc analysis of patients with low plasma lactate dehydrogenase (LDH) levels showed evidence of antitumor activity (*42*). Low LDH reflects a higher cellular dependency on mitochondrial metabolism (as opposed to glycolysis), consistent with our finding that cells undergoing mitochondrial respiration are particularly sensitive to copper ionophores and that this sensitivity is explained by their high levels of lipoylated TCA enzymes. Moreover, our observations that the abundance of FDX1 and lipoylated proteins is highly correlated across a diversity of human tumors and that cell lines with high levels of lipoylated proteins are sensitive to copper-induced cell death suggest that copper ionophore treatment should be directed toward tumors with such a metabolic profile. Future clinical trials of copper ionophores using

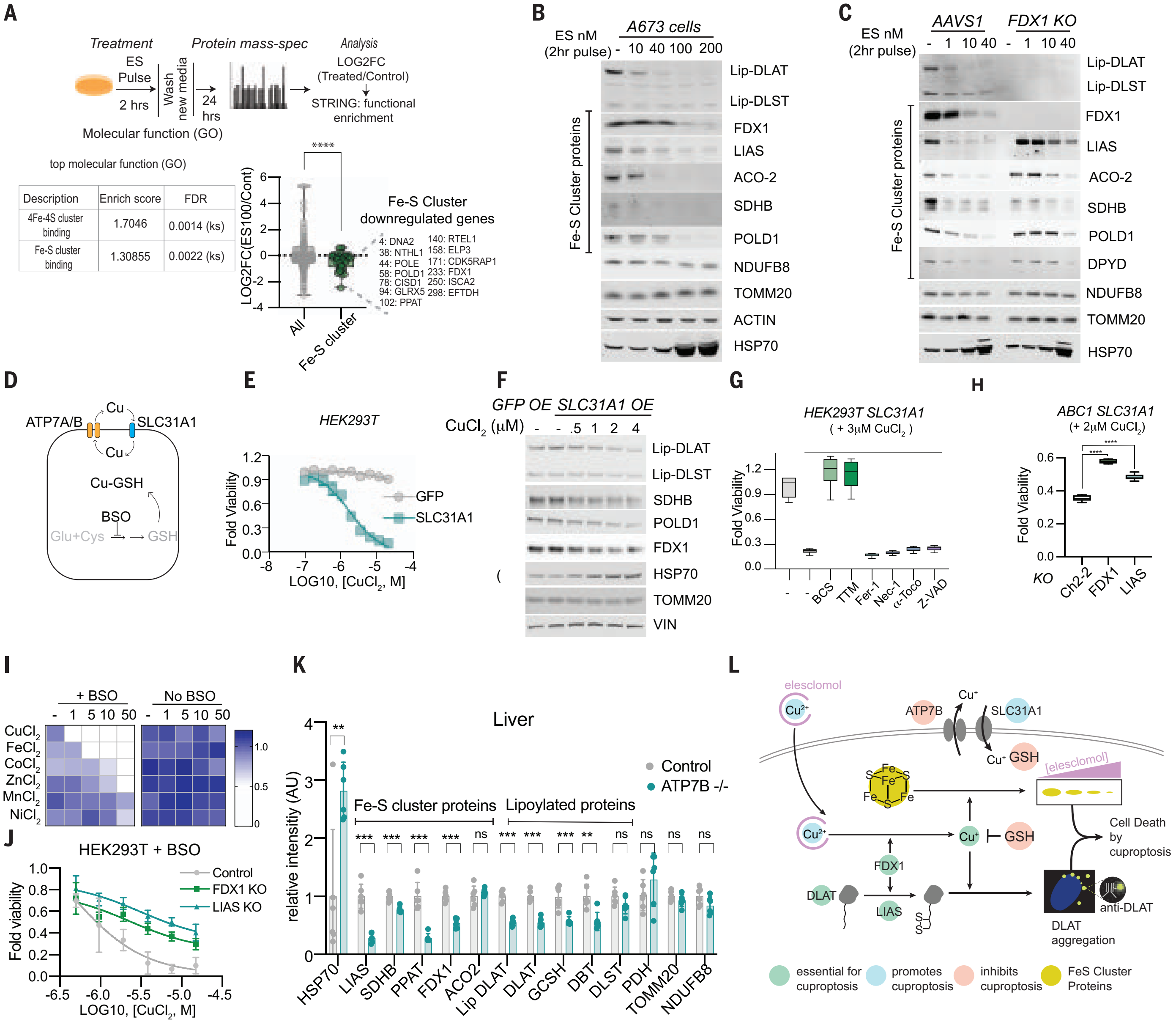


Fig. 6. Shared mechanisms in chemically and genetically induced copper-dependent cell death. (**A**) Proteomic analysis of control and elesclomol (100 nM) pulse-treated ABC1 cells 24 hours after treatment. Top [gene ontology (GO)] enriched categories are presented (table S4). (**B** and **C**) Protein content in A673 cells (B) and AAVS1 and FDX1 KO ABC1 cells (C) 16 hours after 2-hour pulse treatment with elesclomol. For (A) to (C), media were supplemented with 1 μM $CuCl_2$. (**D**) Copper homeostasis schematic. (**E** and **F**) HEK293T cells overexpressing green fluorescent protein (GFP) or SLC31A1 were analyzed for viability 72 hours after supplementation with $CuCl_2$ (E) and for protein content 24 hours after treatment (F). (**G**) Cells overexpressing SLC31A1 were pretreated with 10 μM necrostatin-1, 10 μM ferrostatin-1, 100 μM α-tocopherol, 40 μM Z-VAD-FMK, 10 μM TTM, or 40 μM bathocuproinedisulfonic acid (BCS). Viability was measured after 72 hours of cells growing in the presence or absence of 3 μM $CuCl_2$. (**H**) The viability of ABC1 Ch2-2, *FDX1*, and *LIAS* KO cells overexpressing SLC31A1 was analyzed 72 hours after treatment with 2 μM $CuCl_2$. In (G) and (H), viability is presented as box-and-whisker plots representing interquartile ranges (boxes), medians (horizontal lines), and range (whiskers). For both, $n \geq 5$, and an ordinary one-way analysis of variance (ANOVA) with multiple comparisons was conducted (****adjusted $p < 0.0001$). (**I**) Viability of control and 100 μM BSO–pretreated A549 cells 48 hours after treatment with the indicated metals. An average of three replicates is plotted. The color scale represents the relative viability. (**J**) Viability of control, *FDX1* KO, and *LIAS* KO HEK293T cells pretreated with 20 μM BSO and then with $CuCl_2$. (**K**) Protein content in $Atpb7b^{-/-}$ mouse livers (≥5 $Atpb7b^{-/-}$ mice) compared with control (six $Atpb7b^{+/-}$ mice and one wild-type mouse). Multiple unpaired *t* test analysis was conducted; ***$q < 0.001$, **$q < 0.01$, and ns is not significant. AU, arbitrary units. (**L**) Schematic of mechanisms that promote copper-induced cell death. For (E) and (J), data are means ± SD, with $n \geq 4$.

a biomarker-driven approach should therefore be considered.

REFERENCES AND NOTES

1. B. E. Kim, T. Nevitt, D. J. Thiele, *Nat. Chem. Biol.* **4**, 176–185 (2008).
2. E. J. Ge *et al.*, *Nat. Rev. Cancer* **22**, 102–113 (2022).
3. S. Lutsenko, *Curr. Opin. Chem. Biol.* **14**, 211–217 (2010).
4. T. D. Rae, P. J. Schmidt, R. A. Pufahl, V. C. Culotta, T. V. O'Halloran, *Science* **284**, 805–808 (1999).
5. P. Tsvetkov *et al.*, *Nat. Chem. Biol.* **15**, 681–689 (2019).
6. B. B. Hasinoff, A. A. Yadav, D. Patel, X. Wu, *J. Inorg. Biochem.* **137**, 22–30 (2014).
7. S. Tardito *et al.*, *J. Am. Chem. Soc.* **133**, 6235–6242 (2011).
8. E. W. Hunsaker, K. J. Franz, *Inorg. Chem.* **58**, 13528–13545 (2019).
9. V. Oliveri, *Coord. Chem. Rev.* **422**, 213474 (2020).
10. S. M. Corsello *et al.*, *Nat. Cancer* **1**, 235–248 (2020).
11. D. Cen, D. Brayton, B. Shahandeh, F. L. Meyskens Jr., P. J. Farmer, *J. Med. Chem.* **47**, 6914–6920 (2004).
12. J. R. Kirshner *et al.*, *Mol. Cancer Ther.* **7**, 2319–2327 (2008).
13. S. Tardito *et al.*, *J. Med. Chem.* **55**, 10448–10459 (2012).
14. M. Nagai *et al.*, *Free Radic. Biol. Med.* **52**, 2142–2150 (2012).
15. K. Shimada *et al.*, *Cell Chem. Biol.* **25**, 585–594.e7 (2018).
16. N. C. Yip *et al.*, *Br. J. Cancer* **104**, 1564–1574 (2011).
17. D. Chen, Q. C. Cui, H. Yang, Q. P. Dou, *Cancer Res.* **66**, 10425–10433 (2006).
18. N. Liu, H. Huang, Q. P. Dou, J. Liu, *Oncoscience* **2**, 457–466 (2015).
19. Z. Skrott *et al.*, *Nature* **552**, 194–199 (2017).
20. D. Tang, R. Kang, T. V. Berghe, P. Vandenabeele, G. Kroemer, *Cell Res.* **29**, 347–364 (2019).
21. B. A. Carneiro, W. S. El-Deiry, *Nat. Rev. Clin. Oncol.* **17**, 395–417 (2020).
22. R. Weinlich, A. Oberst, H. M. Beere, D. R. Green, *Nat. Rev. Mol. Cell Biol.* **18**, 127–136 (2017).
23. T. Bergsbaken, S. L. Fink, B. T. Cookson, *Nat. Rev. Microbiol.* **7**, 99–109 (2009).
24. S. J. Dixon *et al.*, *Cell* **149**, 1060–1072 (2012).
25. S. Elmore, *Toxicol. Pathol.* **35**, 495–516 (2007).
26. A. Solmonson, R. J. DeBerardinis, *J. Biol. Chem.* **293**, 7522–7530 (2018).
27. K. Birsoy *et al.*, *Cell* **162**, 540–551 (2015).
28. N. L. Reeder *et al.*, *Antimicrob. Agents Chemother.* **55**, 5753–5760 (2011).
29. A. A. Yadav, D. Patel, X. Wu, B. B. Hasinoff, *J. Inorg. Biochem.* **126**, 1–6 (2013).
30. E. A. Rowland, C. K. Snowden, I. M. Cristea, *Curr. Opin. Chem. Biol.* **42**, 76–85 (2018).
31. M. Ni *et al.*, *Cell Rep.* **27**, 1376–1386.e6 (2019).
32. J. Smirnova *et al.*, *Sci. Rep.* **8**, 1463 (2018).
33. D. Brancaccio *et al.*, *J. Am. Chem. Soc.* **139**, 719–730 (2017).
34. S. Chillappagari *et al.*, *J. Bacteriol.* **192**, 2512–2524 (2010).
35. L. Macomber, J. A. Imlay, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8344–8349 (2009).
36. T. Nevitt, H. Ohrvik, D. J. Thiele, *Biochim. Biophys. Acta* **1823**, 1580–1593 (2012).
37. S. Lutsenko, *Biochem. Soc. Trans.* **36**, 1233–1238 (2008).
38. A. Muchenditsi *et al.*, *Sci. Rep.* **11**, 5659 (2021).
39. O. Bandmann, K. H. Weiss, S. G. Kaler, *Lancet Neurol.* **14**, 103–113 (2015).
40. E. Gaggelli, H. Kozlowski, D. Valensin, G. Valensin, *Chem. Rev.* **106**, 1995–2044 (2006).
41. S. O'Day *et al.*, *J. Clin. Oncol.* **27**, 5452–5458 (2009).
42. S. J. O'Day *et al.*, *J. Clin. Oncol.* **31**, 1211–1218 (2013).
43. L. Cui *et al.*, *Nat. Biotechnol.* **39**, 357–367 (2021).
44. T. Tsang *et al.*, *Nat. Cell Biol.* **22**, 412–424 (2020).
45. C. I. Davis *et al.*, *Metallomics* **12**, 1995–2008 (2020).
46. D. C. Brady, M. S. Crowe, D. N. Greenberg, C. M. Counter, *Cancer Res.* **77**, 6240–6252 (2017).
47. J. A. Mayr, R. G. Feichtinger, F. Tort, A. Ribes, W. Sperl, *J. Inherit. Metab. Dis.* **37**, 553–563 (2014).
48. J. B. Patteson *et al.*, *Science* **374**, 1005–1009 (2021).
49. N. Raffa *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2015224118 (2021).
50. A. Aggarwal, M. Bhatt, *Tremor Other Hyperkinet. Mov. (N. Y.)* **8**, 525 (2018).
51. D. R. Amici *et al.*, *Life Sci. Alliance* **4**, e202000882 (2020).

ACKNOWLEDGMENTS

We thank J. Markley, H. Adelmann, M. Slabicki, V. Wang, and H. Keys for constructive discussion and help with genetic screens. We thank T. Woo for help with immunohistochemistry and immunofluorescence studies and A. Muchenditsi for help with the Atp7b mice. We thank C. Lewis (Whitehead metabolomics), W. Salmon (Whitehead imaging), and R. Rodrigues (Harvard Medical School–Thermo Fisher Scientific Center for Multiplexed Proteomics proteomic facility) for technical help. We thank M. O'Reilly for help with the design of the figures. We thank the Image and Data Analysis Core at Harvard Medical School for coding support. **Funding:** This work was supported by National Cancer Institute (NCI) grant 1 R35 CA242457-01 (T.R.G.), Novo Holdings (T.R.G. and P.T.), NCI grant K08 CA230220 (S.M.C.), National Institute of General Medical Sciences grant T32-GM007748 (S.C.), research and recruitment funding by Boston Children's Hospital (B.P. and N.K.), NCI grant R01-CA194005, NCI grant U54-CA225088, and the Ludwig Center at Harvard (S.S.). **Author contributions:** P.T. conceptualized the project, conducted experiments, collected data, and analyzed results. M.A. and M.D. assisted with experiments. L.J.-C., J.R., and M.K. assisted with data analysis. R.H. and R.D.S. performed the whole-genome CRISPR-Cas9 screens under supervision from S.M.C. S.C. and S.S. performed the tissue microarray staining scoring and visualization, and A.V. assisted with the microscopy analysis. B.P. and N.K. performed and analyzed the metabolomics experiments. S.L. provided study material and experimental advice. J.K.E. and E.F. provided reagents and experimental advice. T.R.G. supervised the research. P.T. and T.R.G. wrote the manuscript. **Competing interests:** S.M.C. and T.R.G. receive research funding unrelated to this project from Bayer HealthCare and Calico Life Sciences. T.R.G. receives research funding related to this project from Novo Holdings, recently held equity in FORMA Therapeutics, is a consultant to GlaxoSmithKline and Anji Pharmaceuticals, and is a founder of Sherlock Biosciences. S.S. is a consultant for RareCyte, Inc. P.T. and T.R.G. are inventors on the patent application PCT/US21/19871 submitted by the Broad Institute entitled "Method of treating cancer." J.E. is currently an employee of Kojin Therapeutics. The other authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abf0529
Materials and Methods
Figs. S1 to S8
Tables S1 to S4
References (*52*–*63*)
MDAR Reproducibility Checklist

30 September 2020; resubmitted 18 January 2022
Accepted 4 February 2022
10.1126/science.abf0529

REPORTS

METALLURGY

Tracking the sliding of grain boundaries at the atomic scale

Lihua Wang[1]†, Yin Zhang[2]†, Zhi Zeng[2], Hao Zhou[2], Jian He[3], Pan Liu[4], Mingwei Chen[5], Jian Han[6], David J. Srolovitz[7,13], Jiao Teng[8], Yizhong Guo[1], Guo Yang[1], Deli Kong[1], En Ma[9], Yongli Hu[10], Baocai Yin[10], XiaoXu Huang[11], Ze Zhang[1,12]*, Ting Zhu[2]*, Xiaodong Han[1]*

Grain boundaries (GBs) play an important role in the mechanical behavior of polycrystalline materials. Despite decades of investigation, the atomic-scale dynamic processes of GB deformation remain elusive, particularly for the GBs in polycrystals, which are commonly of the asymmetric and general type. We conducted an in situ atomic-resolution study to reveal how sliding-dominant deformation is accomplished at general tilt GBs in platinum bicrystals. We observed either direct atomic-scale sliding along the GB or sliding with atom transfer across the boundary plane. The latter sliding process was mediated by movements of disconnections that enabled the transport of GB atoms, leading to a previously unrecognized mode of coupled GB sliding and atomic plane transfer. These results enable an atomic-scale understanding of how general GBs slide in polycrystalline materials.

Grain boundaries (GBs) are interfaces that separate grains of different crystal orientations (*1*, *2*). A GB slides when the two adjoining grains undergo a relative displacement parallel to the boundary plane. GB sliding, sometimes coupled with GB migration, can strongly affect the inelastic deformation of polycrystalline materials, such as diffusional creep at elevated temperature (*3*, *4*) and plastic deformation at ambient temperature (*5*–*7*). GB sliding has recently received considerable attention in studies of plastic deformation and grain size stabilization in nanocrystalline metals (*7*–*12*). Although GB-mediated deformation has been investigated using a variety of in situ and ex situ experimental techniques (*13*–*18*), the atomic-scale process by which GBs slide remains unclear, largely because of a lack of direct high-resolution, experimental observations. The past several years have seen rapid progress in in situ atomic-resolution experiments (*19*–*23*). However, the atomic-scale processes associated with the deformation of general GBs has not been clearly resolved, mainly because tracking atomic movements during GB deformation is extremely difficult.

Atomistic modeling and simulations have provided most of our current understanding of the atomic-scale events accompanying GB deformation (*4*–*7*). For example, molecular dynamics (MD) simulations have examined the mechanical behavior of special types of coincident site lattice tilt GBs in face-center cubic (FCC) bicrystals (*5*–*7*). These simulations revealed a strong coupling between GB sliding and migration through transformations of structural units between the GB and the adjoining crystal lattices (*4*–*7*), which permit sliding with no change in GB structure. This concept has been generalized to account for general GB bicrystallography and the motion of deformation-carrying line defects in GBs, i.e., disconnections (*6*, *7*). Disconnections are characterized by both a step height and Burgers vector: If the step height is zero, then the disconnection is a pure dislocation, whereas if it has zero Burgers vector, then the disconnection is a pure step. MD simulations also demonstrate that the deformation of general GBs, when mediated by GB disconnection motion at room temperature (*24*–*27*), often leads to coupled GB sliding and migration, whereas pure sliding only occurs at relatively high temperatures (*5*–*7*). Unlike special, high-symmetry GBs (*4*–*7*), GBs in real polycrystals are commonly general types, i.e., low-symmetry bicrystallography, asymmetric GB planes, and non-flat GBs. It remains to be demonstrated whether the aforementioned atomic processes of GB sliding, as derived from highly idealized atomistic modeling and simulations, are directly applicable to general GBs in polycrystals under laboratory loading conditions.

We report in situ atomic-scale observations of sliding-dominated deformation of general high-angle tilt GBs in FCC Pt (bicrystals extracted from polycrystals) during nanomechanical testing (*28*, *29*) (fig. S1) in an aberration-corrected transmission electron microscope (Cs-TEM). The thin-film bicrystal geometry permits large-scale GB sliding without the constraints of the triple junctions present in polycrystals, thus facilitating step-by-step monitoring of GB sliding processes over a large sliding distance. We show a time series of in situ Cs-TEM images that provides a representative example of sliding by ~19 Å of an asymmetric <110>-tilt GB with a misorientation angle of 20.1° between adjoining grains (Fig. 1, A to H). A magnified Cs-TEM image of a segment of the initial GB shows that both the left (G_L) and right (G_R) grains are aligned with the <110> zone axis, such that white spots represent the projection of <110> atom-columns (Fig. 1I). General high-angle tilt GBs are rarely flat on the atomic scale. The commonly occurring asymmetric GBs (which are generally observed in real polycrystals) tend to form atomic-scale facets with atomic-sized steps (on one or both sides of the GB plane). For the GB segment studied (Fig. 1I), one side of the boundary exhibits flat, close-packed {111} facets on the face of grain G_R, whereas the other side involves an atomic-scale corrugation on the face of grain G_L, which is a reflection of the fact that this grain face is a high-index {331} plane of grain G_L. This corrugation can be described as single-atomic-layer height steps separated by {111} facets that are three to four atoms wide.

We characterized the structure of the GB (Fig. 1I) in terms of two sets of ½<110>{111} lattice dislocations residing at the GB that serve to accommodate the lattice misorientation between adjoining grains G_L and G_R across this asymmetric tilt GB. We marked each dislocation as "⊥" at the termination of its half plane. These dislocations often combined to form GB Lomer locks, each of which had a characteristic core structure of a five-membered unit (marked by a pentagon). Our Burgers loop analysis of a representative GB Lomer lock (Fig. 1J) shows its Burgers vector to be ½<110>{001} (fig. S2). During GB sliding, each Lomer lock adjusted its position and structure, frequently dissociating temporarily into two individual ½<110>{111} dislocations that subsequently recombined into a Lomer lock. The large open space in the pentagon core unit of a GB Lomer lock facilitates the transfer of atom-columns across the GB plane.

To clarify the nature of atomic-scale GB sliding, we show high-magnification Cs-TEM images (Fig. 2) of the boxed GB region (Fig. 1A) (see movies S1 and S2 and figs. S3 and S4 for the complete time series). In these images, the in-plane atomic rows in G_R are indexed by numbers to facilitate tracking of changes in alignment between atomic rows in G_R and G_L across the GB due to its large sliding displacement (figs. S5 and S6). Image comparison (Fig. 2, A and B) reveals a series of pure GB sliding events. Before GB sliding (Fig. 2A), a close-packed {111} face of G_R was parallel to the GB plane and in contact with a stepped {331} face of G_L; the former consists

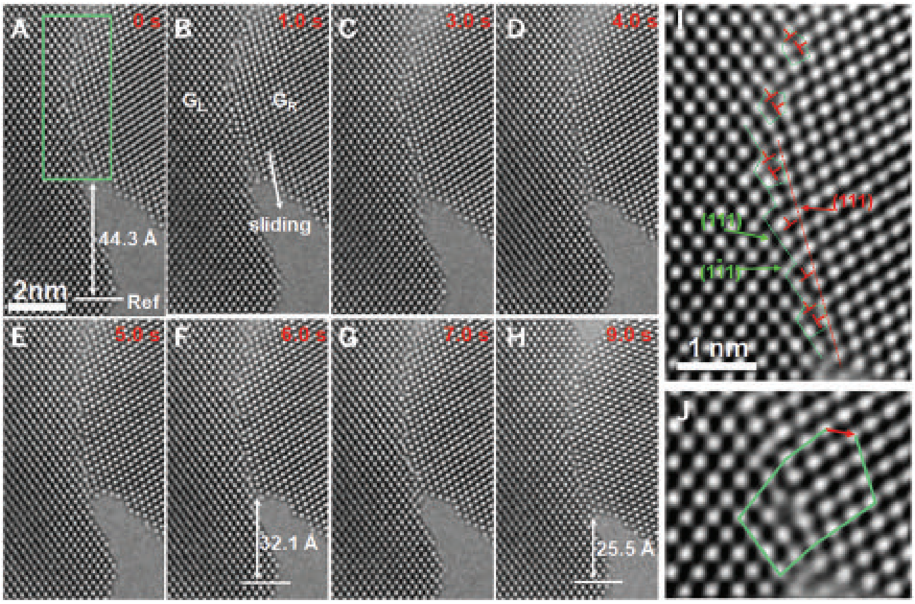


Fig. 1. Atomic-scale sliding of an asymmetric tilt GB in a Pt bicrystal. (**A** to **H**) Series of Cs-TEM images showing the sliding of an asymmetric <110>-tilt GB with 20.1° misorientation. The left and right grains are marked as G_L and G_R, respectively. (**I**) Magnified Cs-TEM image for the green boxed region in (A). One side of the boundary is the close-packed {111} face (marked by a red dashed line) of grain G_R, and the other side consists of a series of atomic-scale steps (marked by green dashed lines) on the face of grain G_L. Two sets of ½<110>{111} GB dislocations are identified, and the extra half plane of each dislocation is marked by symbol "⊥." These GB dislocations often combine to form GB Lomer locks, each of which has the characteristic core structure of a five-membered unit (marked by a pentagon). (**J**) Burgers loop analysis of a representative GB Lomer lock gives the Burgers vector of ½<110>{001} (red arrow).

---

[1]Institute of Microstructure and Property of Advanced Materials, Beijing Key Lab of Microstructure and Property of Advanced Materials, Beijing University of Technology, Beijing 100124, China. [2]Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA. [3]Department of Physics and Astronomy, Clemson University, Clemson, SC 29634 USA. [4]Shanghai Key Laboratory of Advanced High-Temperature Materials and Precision Forming, State Key Laboratory of Metal Matrix Composites, School of Materials Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, P. R. China. [5]Department of Materials Science and Engineering, Johns Hopkins University, Baltimore, MD 21218, USA. [6]Department of Materials Science and Engineering, City University of Hong Kong, Kowloon, Hong Kong SAR, China. [7]Department of Mechanical Engineering, The University of Hong Kong, Hong Kong SAR, China. [8]Department of Material Physics and Chemistry, University of Science and Technology Beijing, Beijing 100083, China. [9]Center for Alloy Innovation and Design (CAID), State Key Laboratory for Mechanical Behavior of Materials, Xi'an Jiaotong University, Xi'an 710049, China. [10]Beijing Institute of Artificial Intelligence, Faculty of Information Technology, Beijing Key Laboratory of Multimedia and Intelligent Software Technology, Beijing University of Technology, Beijing 100124, China. [11]College of Materials Science and Engineering, Chongqing University, Chongqing 40044, China. [12]Department of Materials Science, Zhejiang University, Hangzhou 310008, China. [13]International Digital Economy Academy (IDEA), Shenzhen, China.
***Corresponding author. Email: xdhan@bjut.edu.cn (X.D.H.); ting.zhu@me.gatech.edu (T.Z.); zezhang@zju.edu.cn (Z.Z.)**
†These authors contributed equally to this work.

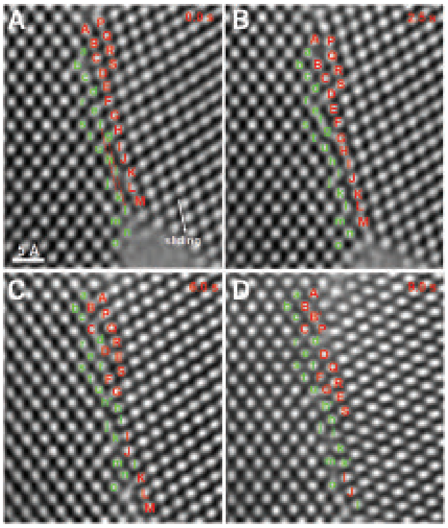


Fig. 2. Atomically resolved sliding at an asymmetric <110>-tilt GB. (**A** to **D**) Magnified Cs-TEM images (corresponding to the green boxed region in Fig. 1A) at 0, 2.5, 6.0, and 9.0 s, showing direct sliding along the boundary plane as well as sliding with atom-column transfer across the boundary plane resulting in coupled GB sliding and atomic plane transfer. At 0 s in (A), atom-columns on the face of grain G_L are marked by green lowercase letters, and atom-columns on the face of grain G_R by red uppercase letters. Two {331} layers on the face of grain G_L are indicated by red dashed lines and they are in contact with the close-packed {111} layer on the G_R face. At 9 s in (D), the two {331} layers were removed from the G_L face; the constituent atom-columns were transferred into the first {111} layer on the G_R face such that they were mixed with atom-columns therein.

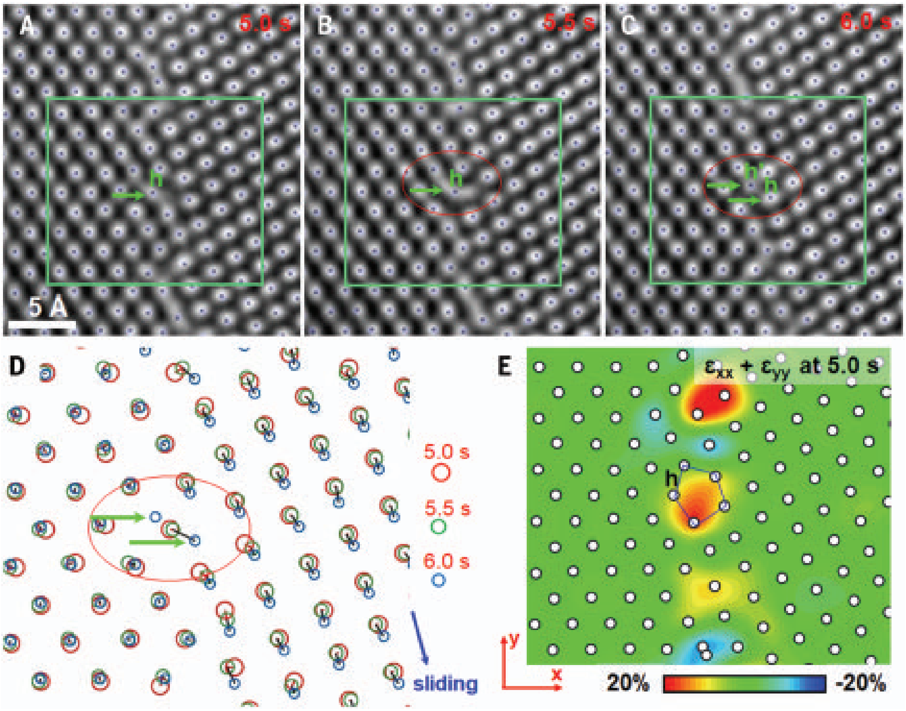


Fig. 3. Automated tracking of atom-column trajectories. (**A** to **C**) Cs-TEM images showing the transfer of atom-column h from the face of grain G_L to the close-packed face of grain G_R, accompanied by the appearance of a new atom-column labeled h′ in (C). Black dots in these images correspond to atom-column centers identified from automated analysis. (**D**) Atom-column trajectories and associated atomic displacement map extracted from (A) to (C) in the green boxed region from automated analysis. Atom-columns at different times are represented by circles of different sizes and colors. Each short segment represents the displacement vector of an atom-column between two consecutive images in (A) to (C), and the length of each segment represents the magnitude of the displacement. The green arrows indicate the transferred atom-column h and the newly formed atom-column h′, respectively. (**E**) Contour map of in-plane atomic volume strains $\varepsilon_{xx} + \varepsilon_{yy}$ at 5 s, calculated in reference to atom-column positions at 4.5 s.

of atom-columns labeled uppercase letters A to M, and the latter comprises atom-columns labeled lowercase letters a to o.

At 0 s, the step-corner atom-column i on the face of G_L was located at a bridge site relative to atom-columns J and K on the face of G_R (Fig. 2A). At 2.5 s, grain G_L slid such that this atom-column i translated to a neighboring bridge site (between columns I and J, Fig. 2B). Other atom-columns (lowercase letters) in the vicinity of column i also slid in a manner similar to that of rigid-body translation. The coherency of slip can also be seen by focusing on another step-corner atom-column g at a bridge site relative to columns H and I at 0 s, which translated to a bridge site relative to columns F and G at 2.5 s. Overall, atom-columns on the face of G_L slid in a rigid body motion relative to those on the face of G_R without rearranging the atomic structure of the faces of G_L and G_R (see fig. S6 for more details). This process was pure (or direct) GB sliding in the absence of GB migration. During this process, the pentagon core units of the GB Lomer locks adjusted their position and/or shape, with no discernable change in misorientation between adjoining grains G_L and G_R. Such pure sliding was mediated by disconnections that glide along the boundary plane (*4*).

In addition to pure GB sliding, we observed a previously unrecognized mode of coupled GB sliding and atomic plane transfer beyond 2.5 s. This GB deformation featured a change of the number of the {331} layers on the face of grain G_L on one side of the GB, which was a result of transfer of the constituent atom-columns of the {331} layers into the first {111} layer on the face of grain G_R on the other side of the GB, in conjunction with transport of transferred atom-columns along the boundary plane. Specifically, comparing GB atom-columns (uppercase and lowercase letters on grains G_L and G_R) at 0 s (Fig. 2A) and those (with mixed uppercase and lowercase letters on the face of grain G_R) at 6.0 and 9.0 s (Fig. 2, C and D), it can be seen that the transfer of atom-columns occurred predominantly from grain G_L to G_R. Focusing on a GB segment at 0 s (Fig. 2A), where the face of grain G_L consisted of the first-layer atom-columns g, i, and l and the second-layer columns f, h, and k, we see that these two atomic layers were closely spaced {331} planes (indicated by red dashed lines). At 6.0 and 9.0 s (Fig. 2, C and D), the constituent atom-columns in the two {331} layers were transferred across the boundary plane and inserted into the first {111} layer on the face of grain G_R (*29*) (fig. S7). This process is evident by the mixed uppercase and lowercase letters in Fig. 2D. Correspondingly, the two {331} layers on the face of G_L present at 0 s were removed by 9 s, thus reducing the number of lattice planes parallel to the GB plane and resulting in a normal displacement of grain G_L by two {331} interlayer spacings toward the GB plane. The coupled GB sliding and atomic plane transfer contrast with the commonly studied GB sliding deformation in previous experiments and simulations in which the number of lattice planes parallel to the GB plane was conserved.

To resolve the unit events of GB sliding more clearly than is possible through visual inspection, we developed an automated atom-column tracking method based on previous

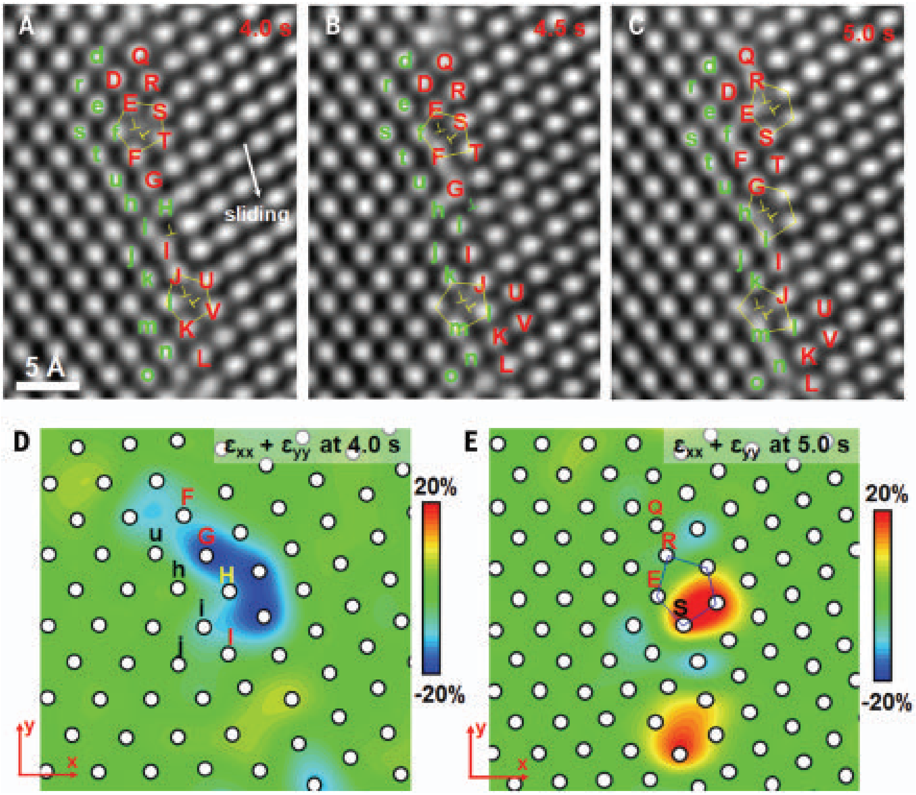


Fig. 4. Movement and formation of GB Lomer locks. (**A**) Two GB Lomer locks were identified from their respective core units (marked by pentagons), along with a stand-alone disconnection in between the two Lomer locks. (**B** and **C**) Movement of the top GB Lomer lock into the face of grain G_R occurred through the formation of a vacant lattice site between atom-columns R and S in (C). A new GB Lomer lock with a pentagon core unit formed in (C) by combining the stand-alone disconnection (marked by unpaired "⊥") in (A) and (B) and a disconnection gliding into the GB segment in the viewport. (**D** and **E**) Contour map of in-plane atomic volume strains $\varepsilon_{xx} + \varepsilon_{yy}$ at 4 and 5 s, respectively.

related studies (*29*–*32*). This method can automatically label atom-columns and thus correlate them between images during GB sliding (fig. S8 and movie S2). To demonstrate its efficacy, we first focused on cases in which the number of atom-columns was not conserved after atom-column transfer across the GB. For example, atom-column H was present at 2.5 s (Fig. 2B) but disappeared at 6 s (Fig. 2C); h′ appeared at 6 s but did not exist at 2.5 s. We were able to resolve the detailed formation process of h′ (Fig. 3) through automated atom-column tracking. At 5 s (Fig. 3A), atom-column h was located at the pentagon core unit of a GB Lomer lock. This core unit enclosed a relatively large free space, which can be considered a vacant lattice site on the face of grain G_R. At 5.5 s (Fig. 3B), a white streak appeared at h during its transfer. This streak was likely caused by the transient processes of transfer of individual atoms that belonged to the same atom-column h, but experienced nonuniform in-plane displacements due to kink nucleation and migration along the atom-column direction (*22*). At 6.0 s (Fig. 3C), h was transferred into the vacant site; meanwhile, a new atom-column h′ appeared at the site occupied previously by the transferred h. Although the total number of atoms should be conserved during GB sliding, the number of atom-columns was increased during the transfer of h (fig. S9). The formation of h′ was further confirmed by the automated tracking of atom-column trajectories, as shown by an atom-column displacement map (Fig. 3D). Atoms in the newly formed atom-column h′ might come from atom-column h and/or its surrounding atom-columns through atom diffusion, but we could not determine this unambiguously from the Cs-TEM images. The diffusive processes of atom transport are likely mediated by vacancy migration. We obtained the map of atomic volume strains at 5 s (Fig. 3E) (*29*). The pentagon core unit of the GB Lomer lock exhibits a large local volume expansion, indicative of a sliding-induced increase of local expansion at this core unit that facilitates the transfer of h. This result implies that the appearance of new atom-columns or the disappearance of preexisting atom-columns (figs. S6 to S11) can effectively accommodate local deformation incompatibilities associated with stepped grain faces, thereby preventing the formation of energetically costly void formation or atom-column jamming at the GB. Therefore, the pentagon core unit of the GB Lomer lock plays an important role in the transfer of atom-columns and the associated formation and removal of atom-columns in the GB region.

To understand how the transfer of two {331} layers on the G_L face occurred through the transfer of constituent atom-columns during GB sliding, we identified a series of unit processes of atom-column transfer across the boundary plane with the aid of GB Lomer locks and subsequent atom-column transport along the GB (*29*) (Fig. 4 and fig. S12). As an example, we focused on the transfer of atom-column E from the face of G_L to G_R. At 4 s (Fig. 4A), we observed two GB Lomer locks in the form of pentagon core units: The top unit was E-f-F-S-T and the bottom was J-l-K-U-V. As GB sliding proceeded, grain G_R moved downward relative to grain G_L. Comparison of Fig. 4, B and C, shows that a vacant lattice site formed between atom-columns R and S such that the top pentagon core unit moved into the first {111} layer of the G_R face. This was followed by the transfer of atom-column E into the newly formed vacant site (fig. S12, A to D), resulting in the motion of the top pentagon core unit back to the G_L face. Although atom-column E left the G_L face, it did not disappear and was actually transferred to the first {111} layer on the G_R face. During the subsequent sliding of grain G_R, atom-column E was moved downward (fig. S12, D to F), resulting in its transport along the GB. We observed similar processes from the bottom pentagon core unit. In this case, atom-column l from the G_L face was first transferred into the first {111} layer of the G_R face (Fig. 4B) and then transported (Fig. 4C and fig. S12, D to F) in the same manner as atom-column E with the aid of the bottom pentagon core unit. As this type of atom-column transfer and motion repeated, several atom-columns were relocated from the {331} layers on the G_L face to the {111} layer on the G_R face (one by one) and then transported down the GB. These processes collectively resulted in the transfer of two {331} layers from the G_L face, leading to the mixing of atom-columns labeled by uppercase and lowercase letters on the G_R face (compare Fig. 2, A and D). Evidently, the coupled processes of GB sliding and atomic plane transfer occurred through a series of displacive events of atom-column transfer and motion.

Comparison of Fig. 4, B and C, reveals that disconnections glided along the GB. Such disconnections were short-lived during GB sliding because of the high stresses applied. Nonetheless, we captured a disconnection (Fig. 4C) due to its combination with a preexisting (intrinsic) GB dislocations to form a new GB Lomer lock in between the top and bottom GB Lomer locks. Subsequent GB sliding led to the annihilation of this temporary lock such that the pentagon core unit decomposed and the preexisting GB dislocation (marked by "⊥") reappeared (fig. S12). This kind of GB Lomer lock formation and decomposition was associated with the displacive process of disconnection gliding and was observed frequently (figs. S11 and S12). In addition, from the map of atomic volume strains at 4 s (Fig. 4D) and 5 s (Fig. 4E), we found a large local volume compression around atom-column H (Fig. 4D) that promoted its removal (Fig. 4B). A large local volume expansion around atom-column s (Fig. 4D) resulted from the transfer of the top pentagon core unit (Fig. 4C). These results, together with those shown in Fig. 3E, underscore the important effects of large local volume strain on driving the formation, removal, and transfer of atom-columns at the GB. We reiterate that the coupled processes of GB sliding and atomic plane transfer occurred through a series of displacive atomic events at ambient temperature and were driven by the applied high stresses at the GB. This stands in contrast to conventional creep deformation of polycrystalline materials operated at elevated temperatures and low stresses, which can be attributed to long-range diffusion of atoms near or at GBs.

Our in situ experiments show that the above atomic-scale processes of GB deformation were also frequently observed at the pentagon core unit of GB Lomer locks from other asymmetric, <110>-tilt GBs with misorientation angles of 82.1° and of 82.7° (figs. S13 to S19). For GBs in polycrystals, the same type of GB structures are common; namely, one side of the GB is a close-packed plane in one grain, and the face of the abutting grain consists of a series of atomic-scale steps. Therefore, the general GB structures and their sliding processes that we observed are representative of a broad class of GBs in polycrystals. In addition, we observed similar atom-column processes during plastic deformation of the mixed type of twist and tilt GBs (fig. S20).

To evaluate the effect of applied loads on driving GB sliding (fig. S1), we conducted peak-pair analysis to determine (elastic) lattice strains near the GB. From the peak-pair analysis strain maps (fig. S21), we calculated the average shear strain (γ_{xy}) near the GB as ~1.5%. The corresponding shear stress was ~980 MPa with the shear modulus of 65.2 GPa for Pt (*33*). Such high shear stress is a substantial fraction of the ideal shear strength (~2.1 GPa) of a perfect Pt crystal (*34*). The average compressive normal strain near the GB was ~0.5% and the corresponding compressive stress was ~850 MPa with an elastic modulus of 170 GPa. The high shear stresses applied enabled GB sliding at room temperature (*35*), and the high compressive stresses drove the transfer of two {331} layers. These high-stress conditions are usually attainable in bulk nanocrystalline and ultra-fine-grained metals, where GB sliding can play an important role in plastic flow (*11*, *15*, *16*). Therefore, the atomic-scale GB deformation processes that we observed should be applicable to bulk nanocrystalline and ultra-fine-grained FCC metals.

Past research has revealed pure GB sliding as well as coupled GB sliding and migration (*1*–*14*). These processes preserve the number of lattice planes parallel to the GB plane. Our in situ atomic-resolution results not only confirm these known sliding modes, but also suggest the possibility of atomic plane transfer during GB sliding, leading to a change of the number of lattice planes parallel to the GB plane. The latter mode can be interpreted in terms of GB disconnection–mediated boundary deformation. A GB disconnection is generally associated with a Burgers vector **b** and a step height h (*7*). During GB sliding from 0 to 2.5s, the GB disconnections involved had a Burgers vector **b** along the GB plane without step component (fig. S22), resulting in pure GB sliding. During coupled GB sliding and atomic plane transfer, the GB disconnections involved had a Burgers vector **b** with both nonzero normal and tangential components with respect to the GB plane (fig. S22). The movement of this type of disconnections produced a GB sliding displacement along the GB plane as well as a relative displacement between adjoining grains normal to the GB plane. In addition, we found that GB atom-columns could be transferred from one side to the other and then moved along the boundary plane upon further GB sliding. This type of displacive motion of atom-columns by an appreciable distance is different from the conventional diffusional creep at elevated temperatures by vacancy-mediated long-range atomic diffusion.

The ability to track time-resolved, atomic-scale motion of GBs opens opportunities for gaining deeper insight into the mechanisms of GB sliding and the mechanical behavior of polycrystalline materials. Some of the observed atomistic mechanisms of sliding of general tilt GBs have not been reported previously. The present results clearly show how general GBs slide at low temperatures and with no (or very little) accompanying GB migration. Broadly, this work demonstrates the great potential of harnessing in situ atomic-resolution TEM experiments for understanding interface-mediated deformation and failure mechanisms in polycrystalline materials and provides new opportunities for high-resolution handshaking between experiment and atomistic modeling.

REFERENCES AND NOTES

1. A. P. Sutton, R. W. Balluffi, *Interfaces in Crystalline Materials* (Oxford Univ. Press, 1995).
2. T. Meiners, T. Frolov, R. E. Rudd, G. Dehm, C. H. Liebscher, *Nature* **579**, 375–378 (2020).
3. R. Raj, M. F. Ashby, *Metall. Trans., A, Phys. Metall. Mater. Sci.* **2**, 1113–1127 (1971).
4. M. F. Ashby, *Surf. Sci.* **31**, 498–542 (1972).
5. J. W. Cahn, Y. Mishin, A. Suzuki, *Acta Mater.* **54**, 4953–4975 (2006).
6. Y. Mishin, M. Asta, J. Li, *Acta Mater.* **58**, 1117–1151 (2010).
7. J. Han, S. L. Thomas, D. J. Srolovitz, *Prog. Mater. Sci.* **98**, 386–476 (2018).
8. J. Schiotz, F. D. Di Tolla, K. W. Jacobsen, *Nature* **391**, 561–563 (1998).
9. H. Van Swygenhoven, P. A. Derlet, *Phys. Rev. B Condens. Matter* **64**, 224105 (2001).
10. V. Yamakov, D. Wolf, S. R. Phillpot, A. K. Mukherjee, H. Gleiter, *Nat. Mater.* **3**, 43–47 (2004).
11. T. Zhu, J. Li, *Prog. Mater. Sci.* **55**, 710–757 (2010).
12. T. Chookajorn, H. A. Murdoch, C. A. Schuh, *Science* **337**, 951–954 (2012).
13. Z. Shan *et al.*, *Science* **305**, 654–657 (2004).
14. T. J. Rupert, D. S. Gianola, Y. Gan, K. J. Hemker, *Science* **326**, 1686–1690 (2009).
15. X. W. Gu *et al.*, *Nano Lett.* **12**, 6385–6392 (2012).
16. Z. H. Aitken, D. Jang, C. R. Weinberger, J. R. Greer, *Small* **10**, 100–108 (2014).
17. X. Zhang *et al.*, *Science* **357**, 397–400 (2017).
18. J. Hu, Y. N. Shi, X. Sauvage, G. Sha, K. Lu, *Science* **355**, 1292–1296 (2017).
19. Q. Zhu *et al.*, *Nat. Commun.* **11**, 3100 (2020).
20. M. Jin, A. M. Minor, E. A. Stach, J. W. Morris Jr., *Acta Mater.* **52**, 5381–5387 (2004).
21. P. Y. Huang *et al.*, *Science* **342**, 224–227 (2013).
22. M. L. Bowers, C. Ophus, A. Gautam, F. Lançon, U. Dahmen, *Phys. Rev. Lett.* **116**, 106102 (2016).
23. A. Azizi *et al.*, *Nat. Commun.* **5**, 4867 (2014).
24. H. Van Swygenhoven, J. R. Weertman, *Mater. Today* **9**, 24–31 (2006).
25. D. S. Gianola, D. Farkas, M. Gamarra, M. He, *J. Appl. Phys.* **112**, 124313 (2012).
26. E. R. Homer, S. M. Foiles, E. A. Holm, D. L. Olmsted, *Acta Mater.* **61**, 1048–1060 (2013).
27. S. L. Thomas, K. Chen, J. Han, P. K. Purohit, D. J. Srolovitz, *Nat. Commun.* **8**, 1764 (2017).
28. L. Wang *et al.*, *Nat. Commun.* **8**, 2142 (2017).
29. Materials and methods are available as supplementary materials.
30. J. C. Crocker, D. G. Grier, *J. Colloid Interface Sci.* **179**, 298–310 (1996).
31. I. F. Sbalzarini, P. Koumoutsakos, *J. Struct. Biol.* **151**, 182–195 (2005).
32. C. Ophus, H. I. Rasool, M. Linck, A. Zettl, J. Ciston, *Adv. Struct. Chem. Imaging* **2**, 15 (2017).
33. D. H. Warner, W. A. Curtin, S. Qu, *Nat. Mater.* **6**, 876–881 (2007).
34. M. Cerny, J. Pokluda, *Comput. Mater. Sci.* **44**, 127–130 (2008).
35. N. Combe, F. Mompiou, M. Legros, *Phys. Rev. B* **93**, 024109 (2016).

ACKNOWLEDGMENTS

X.D.H. and L.W. acknowledge support by the Beijing Outstanding Young Scientists Projects (grant BJJWZYJH01201910005018), the Basic Science Center Program for Multiphase Evolution in Hypergravity of the National Natural Science Foundation of China (grant 51988101), the Beijing Natural Science Foundation (grant Z180014), and the Natural Science Foundation of China (grants 51771004, 51988101, and 91860202). **Author contributions:** L.W. and Y.Z. contributed equally to this work. X.D.H., T.Z., and Z.Z. designed the project and guided the research. L.W. conducted the in situ TEM experiments. Y.Z. developed the automated atom-column tracking method. L.W., H.Z., Y.Z., Zh.Z., J.H., G.Y., E.M., T.Z., D.S., X.X.H., X.D.H., and Z.Z. performed the analysis of TEM results. P.L., M.W.C., J.H., D.S., Y.Z.G., G.Y., D.L.K., Y.L.H., and B.C.Y. performed the peak-pair analysis strain mapping. J.T. synthesized the thin films. L.W., T.Z., E.M., D.S., and X.D.H. wrote the manuscript. All authors contributed to discussion of the results. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm2612
Materials and Methods
Figs. S1 to S23
Table S1
References (*36*–*43*)

12 September 2021; accepted 16 February 2022
10.1126/science.abm2612

RADIO ASTRONOMY

Frequency-dependent polarization of repeating fast radio bursts—implications for their origin

Yi Feng[1,2,3], Di Li[1,2,4]*, Yuan-Pei Yang[5], Yongkun Zhang[1,3], Weiwei Zhu[1], Bing Zhang[6,7], Wenbin Lu[8], Pei Wang[1], Shi Dai[9], Ryan S. Lynch[10], Jumei Yao[11], Jinchen Jiang[1,12], Jiarui Niu[1,3], Dejiang Zhou[1,3], Heng Xu[1,12], Chenchen Miao[1,3], Chenhui Niu[1], Lingqi Meng[1,3], Lei Qian[1], Chao-Wei Tsai[1], Bojun Wang[1,12], Mengyao Xue[1], Youling Yue[1], Mao Yuan[1,3], Songbo Zhang[13], Lei Zhang[1]

The polarization of fast radio bursts (FRBs), which are bright astronomical transient phenomena, contains information about their environments. Using wide-band observations with two telescopes, we report polarization measurements of five repeating FRBs and find a trend of lower polarization at lower frequencies. This behavior is modeled as multipath scattering, characterized by a single parameter, σ_{RM}, the rotation measure (RM) scatter. Sources with higher σ_{RM} have higher RM magnitude and scattering time scales. The two sources with the highest σ_{RM}, FRB 20121102A and FRB 20190520B, are associated with compact persistent radio sources. These properties indicate a complex environment near the repeating FRBs, such as a supernova remnant or a pulsar wind nebula, consistent with their having arisen from young stellar populations.

Fast radio bursts (FRBs) are bright millisecond-duration astronomical transient phenomena observed in radio bands (*1*). Some FRBs, such as FRB 20121102A (*2*), are known to repeat, but it remains unclear whether repeating FRBs (hereafter "repeaters") are ubiquitous or uncommon sources. Whether all FRBs repeat on some time scale, or whether repeaters form a separate population from single-burst sources, has implications for the energy source (*3*), host environments (*4*), luminosity function (*5*), apparent and intrinsic event rates (*6*), and predicted survey yields (*7*).

Information about FRB hosts and environments could be obtained from their polarization. Faraday rotation is the phenomenon whereby the polarization position angle of linearly polarized radiation, propagating through a magneto-ionic medium, is rotated as a function of frequency. It is quantified by the rotation measure (RM), which carries information about the local environment and the intervening medium. The substantial and time-varying RM of FRB 20121102A has been interpreted as due to a dynamic magneto-ionic environment (*8*), such as an expanding supernova remnant or a pulsar wind nebula (*9*, *10*). The polarization position angle and fraction of linear and circular polarization could constrain the emission mechanisms (*11*, *12*). For example, a constant polarization position angle across a burst is consistent with either a model invoking a relativistic shock, or emission from the outer magnetosphere of a neutron star (*3*). Alternatively, the varying polarization angle observed in the repeater FRB 20180301A has been interpreted as originating within the magnetosphere of a magnetar (a highly magnetized neutron star) (*13*).

In this study, we analyzed the polarization properties of 21 FRB sources, 9 of which have been confirmed to repeat. We observed five of them—FRB 20121102A, FRB 20190520B, FRB 20190303A, FRB 20190417A, and FRB 20201124A—with the Five-hundred-meter Aperture Spherical radio Telescope (FAST) and the Robert C. Byrd Green Bank Telescope (GBT). The other 16 sources were taken from the literature (table S1). Our goals were to examine their RM, linear polarization fraction, and its dependence on frequency.

The repeater FRB 20121102A has previously been precisely localized to a dwarf galaxy (*2*, *14*). This FRB has a bimodal burst energy distribution, with 1652 independent bursts detected in 59.5 hours spanning 62 days (*15*). Using the same dataset, we searched RM in each of the 1652 bursts and detected no linear polarization, setting a 6% upper limit on the degree of linear polarization at 1.0 to 1.5 GHz (*16*). For comparison, previous observations showed almost 100% linear polarization at 3 to 8 GHz (*8*).

FRB 20190520B is another repeater localized to a persistent radio source (PRS) in another dwarf galaxy (*17*, *18*). An RM search for 75 bursts (*18*) resulted in no detection, with an upper limit of 20% on the degree of linear polarization at 1.0 to 1.5 GHz (*16*). Our follow-up observations with the GBT detected three bursts in the range of 4 to 8 GHz, with an average RM of 2759 rad/m^2 (*16*). Representative polarization pulse profiles and dynamic spectra of this source are shown in Fig. 1C.

FRB 20190303A, FRB 20190417A, and FRB 20201124A were discovered at 400 to 800 MHz (*19*, *20*). We followed up these sources with the FAST 19-beam system at 1.0 to 1.5 GHz. From FRB 20190303A, three bursts were detected with an average RM of −411 rad/m^2, which help localize the source to right ascension (RA) $13^{\mathrm{h}}51^{\mathrm{m}}58^{\mathrm{s}}$, declination (Dec) +48°07′20″ (J2000 equinox) with a circular 2.′6 uncertainty region (*16*). Compared with the earlier lower-frequency measurements (*19*), the RM has changed by ~100 rad/m^2 in 1.5 years. In contrast to the earlier low linear polarization (*19*), the bursts detected with FAST are nearly 100% linearly polarized.

From FRB 20190417A, we detected 23 bursts, five of which have measurable polarization with an average RM of 4681 rad/m^2. The estimated position of FRB 20190417A is (RA, Dec) = ($19^{\mathrm{h}}39^{\mathrm{m}}22^{\mathrm{s}}$, +59°18′58″) (J2000) with circular uncertainty 2.′6 (*16*). FAST follow-up observations of FRB 20201124A resulted in 11 bursts that were sufficiently bright to be used in our analysis (*16*). We also detected nine polarized bursts with the GBT at 720 to 920 MHz, with an average RM of −684 rad/m^2. Both FRB 20190417A and FRB 20201124A have higher linear polarization at higher frequencies.

The average RMs of each FRB are listed in table S2. The time of arrival, central frequency of the burst emission (weighted by the burst signal-to-noise ratio as a function of frequency), intrachannel depolarization f_{depol} (see below and Eq. 1), RM, and degrees of debiased (*16*) linear and circular polarization of each pulse are listed in table S3. Representative polarization pulse profiles and their dynamic spectra are shown in Fig. 1 (*16*).

The observations of these five sources all indicate decreasing linear polarization with decreasing frequencies. To explain this behavior, we consider three possible effects: (i) intrinsic frequency evolution of the linear polarization, (ii) intrachannel depolarization, and (iii) RM scatter.

Intrinsic frequency evolution of linear polarization is already known for pulsars (*21*). The polarization tends to decrease from lower to higher frequencies, which has been attributed

---

[1]National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China. [2]Zhejiang Lab, Hangzhou, Zhejiang 311121, China. [3]School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, China. [4]National Astronomical Observatories, Chinese Academy of Sciences–University of KwaZulu-Natal Computational Astrophysics Centre, University of KwaZulu-Natal, Durban 4000, South Africa. [5]South-Western Institute for Astronomy Research, Yunnan University, Kunming, Yunnan 650504, China. [6]Department of Physics and Astronomy, University of Nevada, Las Vegas, NV 89154, USA. [7]Nevada Center for Astrophysics, University of Nevada, Las Vegas, NV 89154, USA. [8]Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA. [9]School of Science, Western Sydney University, Penrith, NSW 2751, Australia. [10]Green Bank Observatory, Green Bank, WV 24401, USA. [11]Xinjiang Astronomical Observatory, Chinese Academy of Sciences, Urumqi, Xinjiang 830011, China. [12]School of Physics, Peking University, Beijing 100871, China. [13]Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing 210023, China.
*Corresponding author. Email: dili@nao.cas.cn

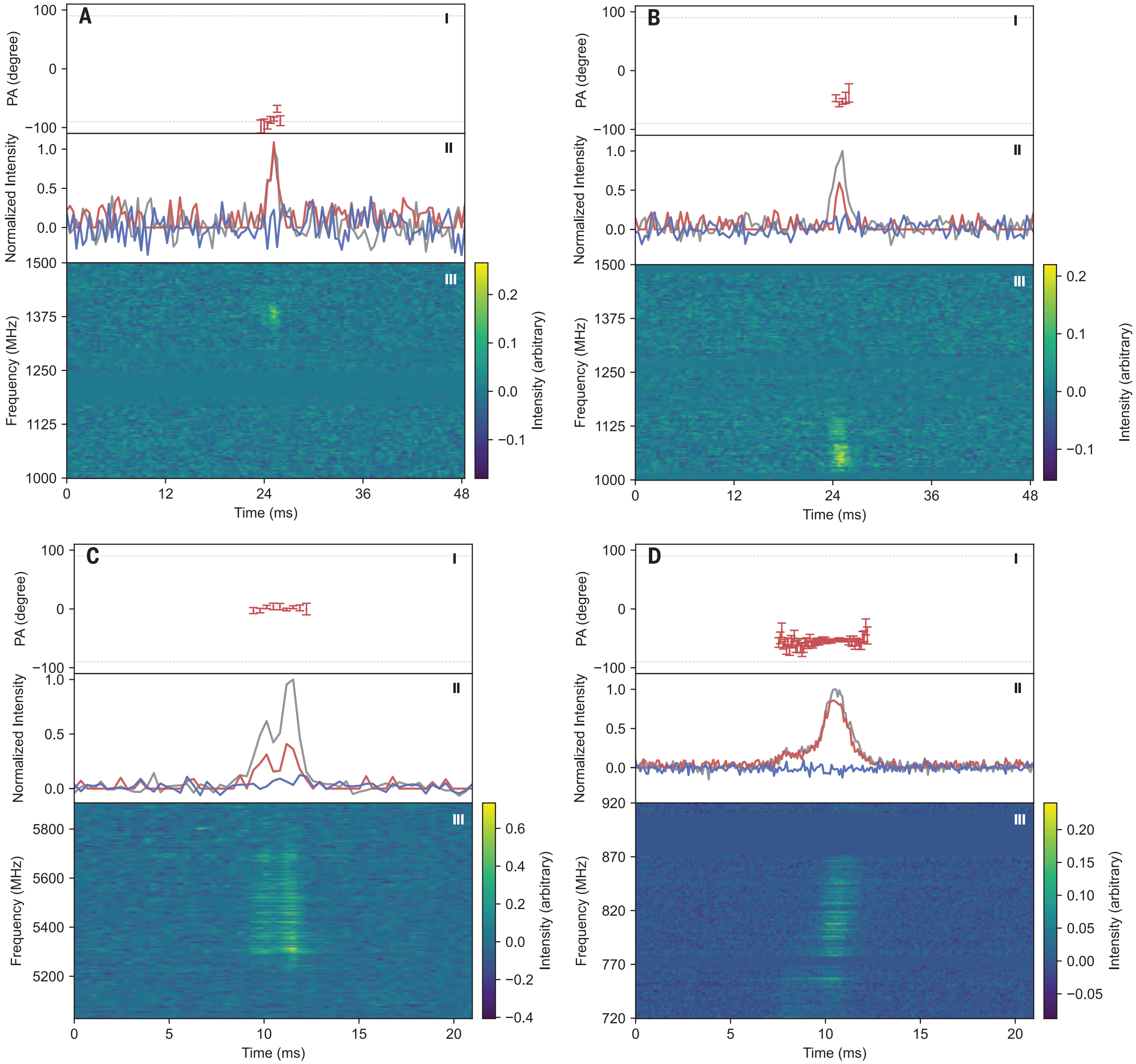


Fig. 1. Polarization position angle and intensity profiles, with dynamic spectra, for four repeating FRBs. (**A**) FRB 20190303A, (**B**) FRB 20190417A, (**C**) FRB 20190520B, and (**D**) FRB 20201124A. In each panel, the top subpanel (I) shows the polarization position angle (PA), with gray dashed lines indicating ±90°; the middle subpanel (II) shows the polarization pulse profiles, with lines indicating total intensity (black, normalized to a peak value of 1.0), linear polarization (red), and circular polarization (blue); and the bottom subpanel (III) shows the dynamic spectra (color bar). Error bars in each subpanel (I) indicate 1σ error for the polarization position angle.

to emission from different heights in the pulsar magnetospheres (*22*). This is the opposite trend to the one we find for repeating FRBs, so a direct analogy is not supported by the data. Given the lack of understanding of FRB origin(s), our results cannot rule out other scenarios involving pulsar-magnetosphere–like environments.

For the alternative intrachannel depolarization, the fractional reduction in the linear polarization amplitude, $f_{\rm depol}$, is defined as

$$f_{\rm depol} \equiv 1 - \frac{\sin(\Delta\theta)}{\Delta\theta} \qquad (1)$$

where $\Delta\theta$ is the intrachannel polarization position angle rotation. $\Delta\theta$ is defined as $\Delta\theta \equiv \frac{2{\rm RM_{obs}}c^2\Delta\nu}{\nu_c^3}$, where $\rm RM_{obs}$ is the observed rotation measure, c is the speed of light, $\Delta\nu$ is the channel frequency width, and ν_c is the central channel observing frequency. For repeaters, the measured RMs are too small to explain the depolarization through this effect ($f_{\rm depol}$ listed in tables S1 and S3). We infer that intrachannel depolarization is unlikely to be a major cause of depolarization for repeating FRBs.

We next examine RM scattering, the dispersion of RM about the apparent mean for each source (*23*). RM scattering can be caused

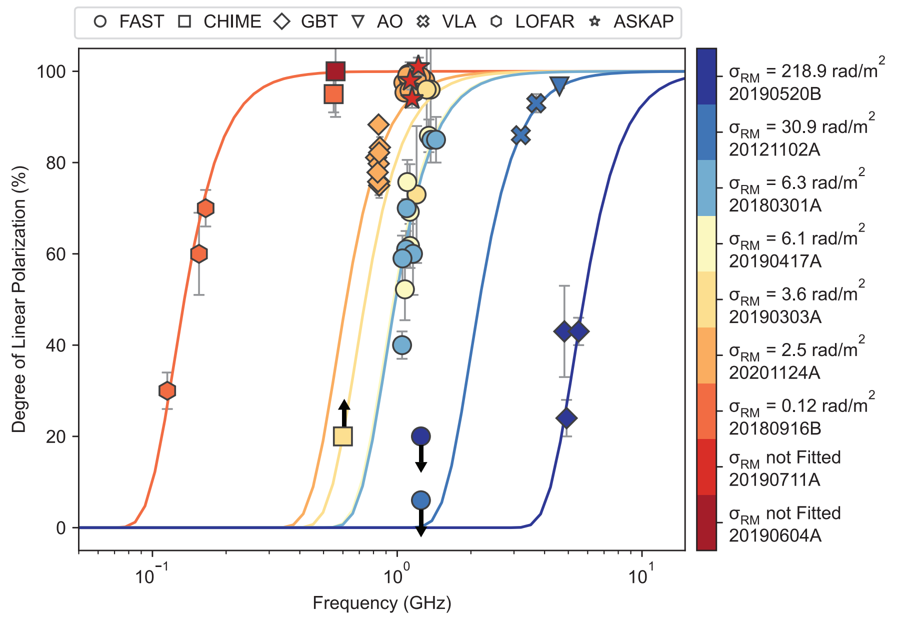


Fig. 2. The degree of linear polarization for FRB sources is consistent with RM scattering. Data points (with 1σ error bars) indicate the degree of linear polarization as a function of frequency for each FRB (indicated at right). The colored lines are models of emission that is intrinsically 100% linearly polarized, then depolarized by various σ_{RM} levels following the model in Eq. 2, fitted to each FRB separately. Arrows indicate 95% upper and lower limits. All bursts in the sample are consistent with an RM scattering model. Data values and sources are listed in tables S1 and S3; fitted σ_{RM} values are in table S2. Symbol shapes indicate the telescope used for each observation: Five-hundred-meter Aperture Spherical radio Telescope (FAST), Canadian Hydrogen Intensity Mapping Experiment (CHIME), Robert C. Byrd Green Bank Telescope (GBT), Arecibo Observatory (AO), Karl G. Jansky Very Large Array (VLA), Low-Frequency Array (LOFAR), and Australian Square Kilometre Array Pathfinder (ASKAP).

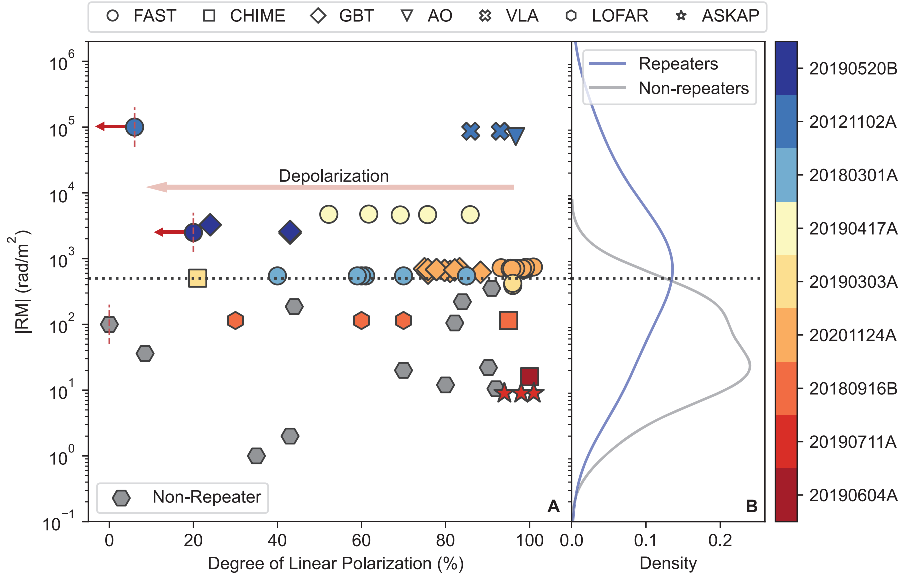


Fig. 3. Relationship between RM and degree of linear polarization. (**A**) Data for repeating (colored points) and nonrepeating FRBs (gray points). Symbol shapes and colors are the same as in Fig. 2. Vertical dashed red lines indicate nondetections of RM in the L band; thus, the associated data points are displayed at nominal values that are either representative or measured at higher frequencies: 100 rad/m^2 for FRB 20140514A, 3000 rad/m^2 for FRB 20190520B, and 10^5 rad/m^2 for FRB 20121102A. Red arrows denote 95% upper limits on the linear polarization. The horizontal dotted black line, which lies above all nonrepeaters, indicates $|RM| = 500$ rad/m^2. (**B**) Kernel density estimation of the RMs for repeaters (blue line) and nonrepeaters (gray line). A Kolmogorov-Smirnov test between the repeaters and nonrepeaters finds a *P* value of 0.02, indicating that they are statistically different from each other.

by multipath transmission of signals in an inhomogeneous magneto-ionic environment. If the scattering is sufficiently large, it can become substantial enough to depolarize the pulses, analogous to pulsar pulses passing through a stellar wind (*24*, *25*). We parameterize the depolarization due to RM scattering as (*23*)

$$f_{\text{RM scattering}} \equiv 1 - \exp\left(-2\lambda^4\sigma_{\text{RM}}^2\right) \qquad (2)$$

where $f_{\text{RM scattering}}$ is the fractional reduction in the linear polarization amplitude, σ_{RM} is the standard deviation of the RM, and λ is the wavelength.

Depolarization at lower frequencies, consistent with irregular RM variations, has been seen in a few pulsars with scatter broadening. For example, the variable degree of linear polarization observed in the pulsar PSR J0742−2822 between 200 MHz and 1 GHz can be described by Eq. 2 with $\sigma_{\text{RM}} = 0.13$ rad/m^2 (*26*).

Complex magneto-ionic environments have previously been inferred from observations for some repeaters (*8*, *13*). We also expect RM scatter for FRBs in such environments. In Fig. 2, we show the degree of linear polarization versus frequency for each source individually. Frequency evolution can be seen for all sources, but the depolarization occurs at different bands. FRB 20180916B is depolarized below 200 MHz (*27*), whereas FRB 20121102A (*28*) is depolarized at frequencies lower than 3.5 GHz. We fitted the data for each repeater using the model in Eq. 2, with a different σ_{RM} for each source (*16*). For the sources depolarized at lower frequencies (<200 MHz), such as FRB 20180916B, we derive $\sigma_{\text{RM}} \sim 0.1$ rad/m^2 (table S2). Such a small scatter is consistent with the environment being less turbulent, dense, and/or magnetized than that of other FRBs, as might be expected for an old stellar population (*27*). We find $\sigma_{\text{RM}} \sim 2.5$ rad/m^2 for FRB 20201124A, $\sigma_{\text{RM}} \sim 3.6$ rad/m^2 for FRB 20190303A, $\sigma_{\text{RM}} \sim 6.1$ rad/m^2 for FRB 20190417A, $\sigma_{\text{RM}} \sim 6.3$ rad/m^2 for FRB 20180301A, and $\sigma_{\text{RM}} \sim 30.9$ rad/m^2 for FRB 20121102A (table S2). The RM scatter of 30.9 rad/m^2 is consistent with FRB 20121102A originating in a younger stellar population (*28*) that produced an inhomogeneous magneto-ionic environment. For the repeater FRB 20190520B, which depolarizes at frequencies higher than 1 GHz, we derive $\sigma_{\text{RM}} \sim 218.9$ rad/m^2.

The observations are consistent with repeating bursts being intrinsically nearly 100% linearly polarized, then being depolarized during propagation, which can be characterized by the RM scatter parameter σ_{RM}. The same environment that gives rise to large σ_{RM} could also cause Faraday conversion, whereby linearly polarized light is converted to circularly polarized light, potentially explaining the circular polarization observed in some FRBs (*29*). We

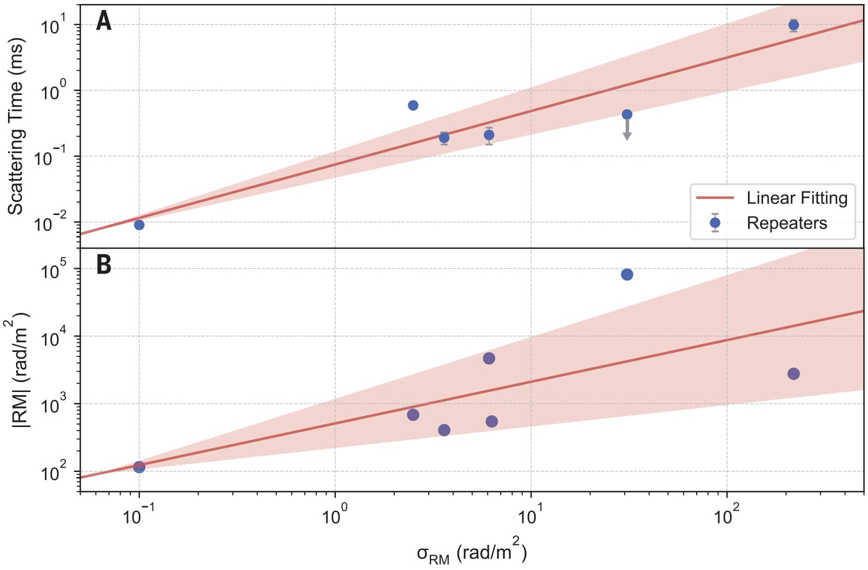


Fig. 4. Correlations between σ_{RM}, scattering time τ_{sca}, and rotation measure magnitude |RM| for repeating FRBs. (**A**) Relationship between σ_{RM} and scattering time scaled to 1300 MHz (*16*) for repeaters with σ_{RM} measurements (error bars, 1σ error for the scattering time). The data used are listed in table S2. FRB 20121102A is an upper limit (gray arrow). The Pearson product-moment correlation coefficient of $\log\tau_{sca}$ and $\log\sigma_{RM}$ is 0.47. Fitting the data with a linear model (red line) yields a slope of 0.81 ± 0.16. (**B**) Relationship between σ_{RM} and |RM| for the same sample of repeaters. The Pearson product-moment correlation coefficient of log|RM| and $\log\sigma_{RM}$ is 0.68. Fitting the data with a linear model (red line) yields a slope of 0.62 ± 0.30. All quantities are measured in the observer's frame. The 1σ uncertainties on σ_{RM} and |RM| are smaller than the symbol size.

interpret the σ_{RM} values derived from our analysis as a measure of the complexity level of the magneto-ionic environments that host active repeaters, with larger σ_{RM} values possibly being associated with younger stellar populations.

Figure 3 shows the relation between RM and linear polarization for the repeating and nonrepeating FRBs in our sample that have polarization measurements. A Kolmogorov-Smirnov test between the repeaters and nonrepeaters shows that the RM distributions differ, indicating that they may reside in different environments. We caution that the intrachannel depolarization may introduce a selection bias if the nonrepeaters have systematically larger RMs, because discovery searches tend to use coarser filter banks than follow-up observations.

If σ_{RM} is due to multipath propagation effects on individual FRBs, we would expect the same effect to produce a temporal scattering in the pulse profile. In Fig. 4A, we show RM scatter and scattering time scales (listed in table S2) for our repeater sample. The two are positively correlated, consistent with the hypothesis that RM scatter and pulse scattering originate from a single plasma screen, such as a supernova remnant or pulsar wind nebula (*30*). σ_{RM} and |RM| are also positively correlated (Fig. 4B). In the context of our multipath-scattering model, this indicates that an environment with a stronger magnetic field strength *B* tends to have a larger fluctuation in *B*.

FRB 20190520B and FRB 20121102A are the only two FRBs known to have associated compact PRSs (*4*, *18*). We find that these two repeaters have the largest σ_{RM}. The combination of large RM (dense) and strong *B* field (magnetized) tends to produce large σ_{RM}. A denser and more magnetized environment likely also produces stronger synchrotron radiation from the nebula, resulting in a PRS (*31*), consistent with the multipath-scattering picture. Repeaters with large observed σ_{RM} could be more affected by turbulence, resulting in large fluctuations of electron density and magnetic field, which may explain the diversity among repeaters (*30*).

In summary, repeating FRBs are less polarized at lower frequencies, which can be explained by an RM scatter model. The value of σ_{RM} in the model can be used to quantify the complexity of magnetized environments associated with repeating FRBs. High values of σ_{RM} possibly indicate sources in younger stellar populations.

REFERENCES AND NOTES

1. D. R. Lorimer, M. Bailes, M. A. McLaughlin, D. J. Narkevic, F. Crawford, *Science* **318**, 777–780 (2007).
2. L. G. Spitler *et al.*, *Nature* **531**, 202–205 (2016).
3. B. Zhang, *Nature* **587**, 45–53 (2020).
4. B. Marcote *et al.*, *Astrophys. J. Lett.* **834**, L8 (2017).
5. R. Luo, K. Lee, D. R. Lorimer, B. Zhang, *Mon. Not. R. Astron. Soc.* **481**, 2320–2337 (2018).
6. D. W. Gardenier, J. van Leeuwen, L. Connor, E. Petroff, *Astron. Astrophys.* **632**, A125 (2019).
7. J.-P. Macquart, R. D. Ekers, *Mon. Not. R. Astron. Soc.* **474**, 1900–1908 (2018).
8. D. Michilli *et al.*, *Nature* **553**, 182–185 (2018).
9. A. L. Piro, B. M. Gaensler, *Astrophys. J.* **861**, 150 (2018).
10. B. Margalit, B. D. Metzger, *Astrophys. J. Lett.* **868**, L4 (2018).
11. W. Lu, P. Kumar, R. Narayan, *Mon. Not. R. Astron. Soc.* **483**, 359–369 (2019).
12. S. Dai *et al.*, *Astrophys. J.* **920**, 46 (2021).
13. R. Luo *et al.*, *Nature* **586**, 693–696 (2020).
14. S. Chatterjee *et al.*, *Nature* **541**, 58–61 (2017).
15. D. Li *et al.*, *Nature* **598**, 267–271 (2021).
16. Materials and methods are available as supplementary materials.
17. D. Li *et al.*, *IEEE Microw. Mag.* **19**, 112–119 (2018).
18. C.-H. Niu *et al.*, arXiv:2110.07418 [astro-ph.HE] (2021).
19. E. Fonseca *et al.*, *Astrophys. J. Lett.* **891**, L6 (2020).
20. CHIME/FRB Collaboration, The Astronomer's Telegram no. 14497 (2021); www.astronomerstelegram.org/?read=14497.
21. S. Johnston, M. Kerr, *Mon. Not. R. Astron. Soc.* **474**, 4629–4636 (2018).
22. S. A. Petrova, *Astron. Astrophys.* **378**, 883–897 (2001).
23. S. P. O'Sullivan *et al.*, *Mon. Not. R. Astron. Soc.* **421**, 3300–3315 (2012).
24. X. P. You, R. N. Manchester, W. A. Coles, G. B. Hobbs, R. Shannon, *Astrophys. J.* **867**, 22 (2018).
25. E. J. Polzin *et al.*, *Mon. Not. R. Astron. Soc.* **490**, 889–908 (2019).
26. M. Xue *et al.*, *Publ. Astron. Soc. Aust.* **36**, e025 (2019).
27. Z. Pleunis *et al.*, *Astrophys. J. Lett.* **911**, L3 (2021).
28. G. H. Hilmarsson *et al.*, *Astrophys. J. Lett.* **908**, L10 (2021).
29. G. H. Hilmarsson, L. G. Spitler, R. A. Main, D. Z. Li, *Mon. Not. R. Astron. Soc.* **508**, 5354–5361 (2021).
30. Detailed modeling and discussions can be found in the supplementary text.
31. Y.-P. Yang, Q.-C. Li, B. Zhang, *Astrophys. J.* **895**, 7 (2020).
32. D. Li, Frequency dependent polarization of repeating fast radio bursts - implications for their origin, version 4, Science Data Bank (2022); https://doi.org/10.11922/sciencedb.o00069.00006.

ACKNOWLEDGMENTS

We thank B. Gregory for help with the observations and E. Fonseca, G. H. Hilmarsson, D. Michilli, J. Hessels, A. Deller, C. Day, S. White, and B. Simon for valuable discussions. This work made use of data from FAST, a Chinese national mega-science facility, operated by National Astronomical Observatories, Chinese Academy of Sciences. The Green Bank Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. **Funding:** D.L. and Y.F. are supported by NSFC grant nos. 11988101 and 11725313; by the National Key R&D Program of China no. 2017YFA0402600; and by Key Research Project of Zhejiang Lab no. 2021PE0AC03. Y.-P.Y. is supported by NSFC grant no. 12003028. W.Z. is supported by National SKA Program of China no. 2020SKA0120200 and NSFC nos. 12041303 and 11873067. W.L. is supported by the Lyman Spitzer, Jr. Fellowship at Princeton University. P.W. is supported by NSFC grant no. U2031117, the Youth Innovation Promotion Association CAS (ID 2021055), CAS Project for Young Scientists in Basic Research (grant no. YSBR-006), and the Cultivation Project for FAST Scientific Payoff and Research Achievement of CAMS-CAS. S.D. is the recipient of an Australian Research Council Discovery Early Career Award (no. DE210101738) funded by the Australian Government. J.Y. is supported by NSFC grant no. 11903049 and the Cultivation Project for FAST Scientific Payoff and Research Achievement of CAMS-CAS. L.Q. is supported by National SKA Program of China no. 2020SKA0120100 and the Youth Innovation Promotion Association of CAS (ID 2018075). S.Z. is supported by NSFC grant no. 12041306 and by China Postdoctoral Science Foundation grant no. 2020M681758. L.Z. is supported by NSFC grant no. 12103069. **Author contributions:** Y.F. and D.L. developed the concept of the study and led the interpretation of the results and writing of the manuscript. Y.-P.Y., B.Z., and W.L. derived the theoretical formulae. Y.F., D.L., W.Z., B.Z., W.L., R.S.L., and C.N. arranged the observations. Y.F. and Y.Z. conducted the main data analysis and visualization.

All authors contributed to the analysis or interpretation of the data and to the final version of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The FAST and GBT data from our observations of the five repeating FRBs are available (in PSRCHIVE format) in the Science Data Bank (*32*). The archival data sources are listed in table S1. Computer code for our polarization analysis and figure plotting is available at https://github.com/SukiYume/RMS. The numerical results of our analysis are listed in tables S2 and S3.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl7759
Materials and Methods
Supplementary Text
Figs. S1 to S5
Tables S1 to S3
References (*33*–*62*)

4 August 2021; accepted 9 February 2022
10.1126/science.abl7759

ORGANIC CHEMISTRY

Concise syntheses of GB22, GB13, and himgaline by cross-coupling and complete reduction

Eleanor M. Landwehr†, Meghan A. Baker†‡, Takuya Oguma†§, Hannah E. Burdge¶, Takahiro Kawajiri§, Ryan A. Shenvi*

Neuroactive metabolites from the bark of *Galbulimima belgraveana* occur in variable distributions among trees and are not easily accessible through chemical synthesis because of elaborate bond networks and dense stereochemistry. Previous syntheses of complex congeners such as himgaline have relied on iterative, stepwise installation of multiple methine stereocenters. We decreased the synthetic burden of himgaline chemical space to nearly one-third of the prior best (7 to 9 versus 19 to 31 steps) by cross-coupling high fraction aromatic building blocks (high F_{sp}2) followed by complete, stereoselective reduction to high fraction sp^3 products (high F_{sp}3). This short entry into *Galbulimima* alkaloid space should facilitate extensive chemical exploration and biological interrogation.

Extracts from the rainforest canopy trees *Galbulimima belgraveana* and *G. baccata* have furnished a variety of neuroactive alkaloids (*1*, *2*) that are classified chemically by connectivity between piperidine and decalin domains [*Galbulimima* (GB) alkaloids, classes I to IV] (Fig. 1, A and B) (*3*). Different connectivity and substitution patterns confer diverse and profound biological effects in mammals (*2*), but only a single target has been defined. The simplest class I GB alkaloid, himbacine [complexity measurement of 15 molecular bits (mcbits)/atom] (*4*), was found to antagonize muscarinic receptors $M_1$to M_5 [rhodopsin-like G protein–coupled receptors (GPCRs), subfamily A18] (*5*, *6*). An analog of *ent*-himbacine (non-natural) was discovered through total synthesis to antagonize the related GPCR protease-activated receptor–1 (PAR-1) (subfamily A15) (*7*) and was developed into Vorapaxar, establishing an important precedent for the perturbation of selectivity among metabotropic receptors with medicinal potential. Whereas class I alkaloids effect bradycardia and putative psychotropic activity in mice, class II alkaloids cause tachycardia and hypotension, and class III alkaloids such as himgaline (**1**) are antispasmodic (class IV alkaloids have not been assayed) (*2*). One or more components of *G. belgraveana* bark cause hallucination in humans, but a psychotropic principle has not been identified (*1*). Aside from himbacine, the remaining alkaloids have not been assigned biological targets, obstructed by the alkaloid content in bark—a mixture of more than 30 congeners that varies unpredictably in abundance from 0.5% to only trace amounts (*8*)—and synthetic intractability (*3*). This compound problem of poor material access, absence of target annotation, and translational potential of multiple bioactive metabolites has attracted substantial interest from the synthesis community (*3*, *9*–*14*). GB alkaloid classes II and III have proved especially challenging to access (18 to 33 steps), although pioneering solutions have appeared (full outlines of each are available in figs. S1 to S7) (*9*–*14*). The most concise syntheses by Movassaghi (*10*), Sarpong (*13*), and Ma (*14*) obtained GB13 [(–)-**2** or (+)-**2**, (±)-**2**, and (–)-**2**, respectively] in 18 to 19 steps and generated crucial data to aid our exploration.

In search of a general strategy to access GB alkaloid chemical space, we targeted GB22 (**3**) (Fig. 1B)—a low-abundance class III alkaloid [1.8 parts per million (ppm)], milled bark (*15*) containing an aromatic ring that might allow rapid entry and diversification to complex congeners. In this design, high fraction aromatic intermediate **5** containing 15 sp^2 atoms (11 mcbits/atom) might be reduced (*13*, *16*) to generate the 11 tetrahedral stereocenters in the complex alkaloid himgaline (**1**; 21 mcbits/atom). This is a classical strategy, represented by the earliest alkaloid synthesis—coniine in 1886—in which reduction converted a simple aromatic to a high fraction sp^3 product (high F_{sp}3) alkaloid, albeit with one stereocenter, the stereochemistry of which was not controlled (*17*). Success in accessing himgaline would require relay of stereochemical configuration from the single stereocenter of **5** to nine prochiral carbons, including the concave-facing methine C–H bonds of GB13 (**2**). Although these hydrogens might epimerize to the concave face, the potential instability of extended enols and the existence of other low-energy configurations at ring junctions complicated our analysis of stereochemical equilibration [for example, 9,10,15-*epi*-GB13 and 15-*epi*-GB13 are 2.6 kcal/mol and 1.3 kcal/mol lower in energy, respectively, than GB13 (MM2) (figs. S8–10)]. Here, we report unusually concise syntheses of GB22, GB13, and himgaline by using an *endo*-selective attached-ring cross-coupling and arene reduction that significantly reduce the synthetic burden compared with that of prior art.

Retrosynthetic analysis of GB22 (Fig. 1B) identified an embedded attached-ring system, which could be unmasked by pyridine hydrogenation (*13*) and intramolecular ketone arylation transforms. Despite the simplicity of **5**, the most obvious disconnection—enone conjugate addition—fails. A direct Friedel-Crafts addition of **4** into the cyclohexadienone conjugate acid of **3** (Fig. 2A) (*18*) is prevented by preferential protonation of **4** and decomposition of **5** mediated by acids. We thought that if the oxocarbenium ion were replaced with a β-keto carbon–centered radical, we might circumvent the Friedel-Crafts through interception of an arylnickel complex (Fig. 2C) (*19*). β-Keto radical formation has been implicated in the ring-opening of siloxycyclopropanes with photoinduced electron transfer (PET) to 1,4-dicyanonaphthalene (DCN) (*20*). Inspired by the recent success of photoredox-Ni dual catalytic cross-coupling platforms (*21*), we considered a system in which a photoexcited catalyst might oxidatively cleave a siloxycyclopropane with *endo*-selectivity (*22*–*24*), leading to arylnickel capture and reductive elimination. Typically, transition metal–catalyzed arylations of cyclopropanols and siloxycyclopropanols favor cleavage of the least-hindered cyclopropane bond (Fig. 1C, path a) (*25*). By contrast, this electron transfer arylation would provide the opposite regioselectivity (Fig. 1C, path b).

Early efforts to develop a dual catalytic *endo*-selective siloxycyclopropane arylation identified the Doyle-MacMillan dual-catalyst system of $\{Ir[dF(CF_3)ppy]_2(dtbbpy)\}PF_6$ and Ni(dtbbpy)Cl_2 as a good starting point (Fig. 2D)

Department of Chemistry, Scripps Research, La Jolla, CA, USA.
***Corresponding author. Email: rshenvi@scripps.edu**
†These authors contributed equally to this work.
‡Present address: Merck & Co., South San Francisco, CA, USA.
§Present address: Shionogi & Co., Toyonaka, Japan.
¶Present address: Pfizer, La Jolla, CA, USA.

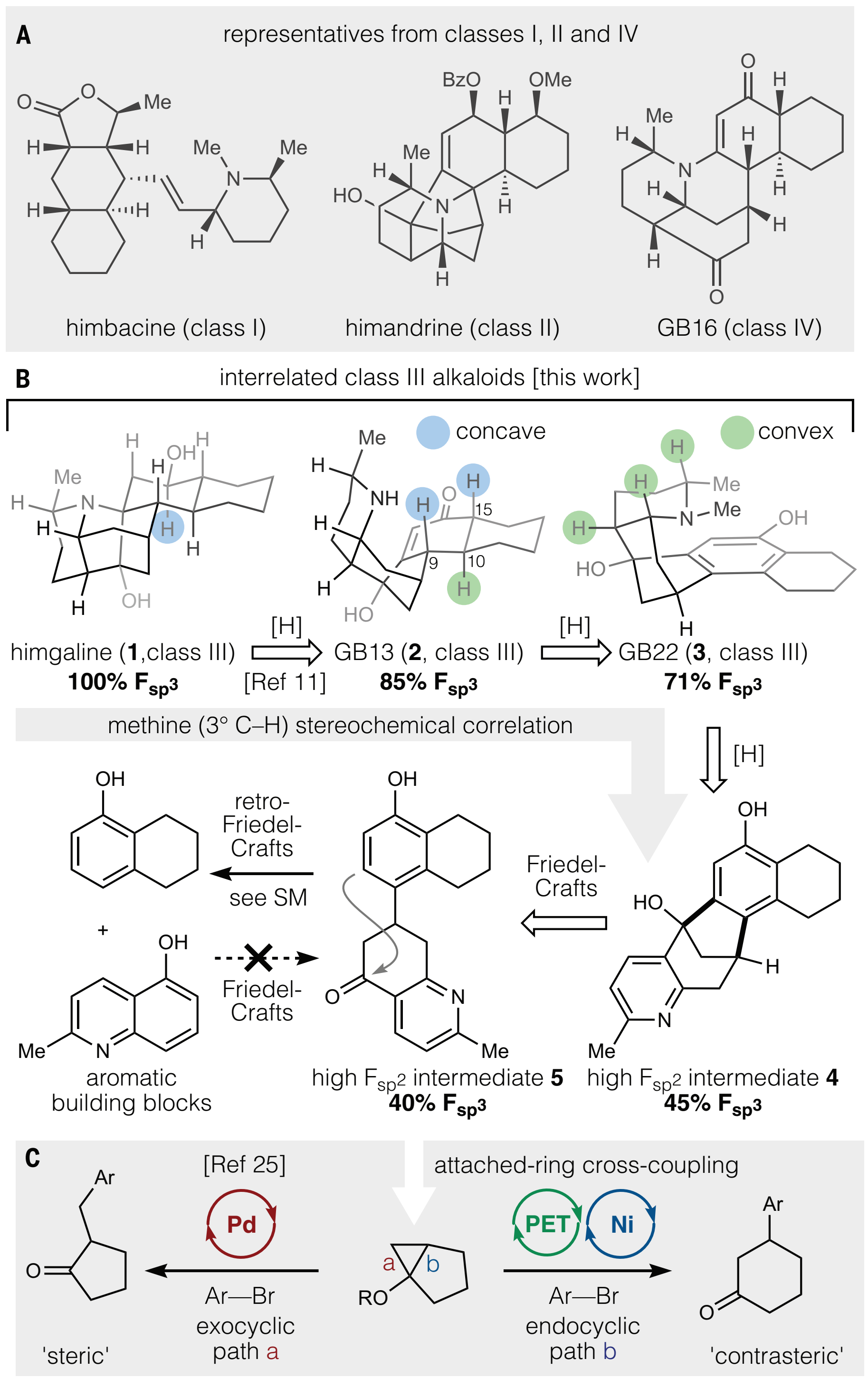
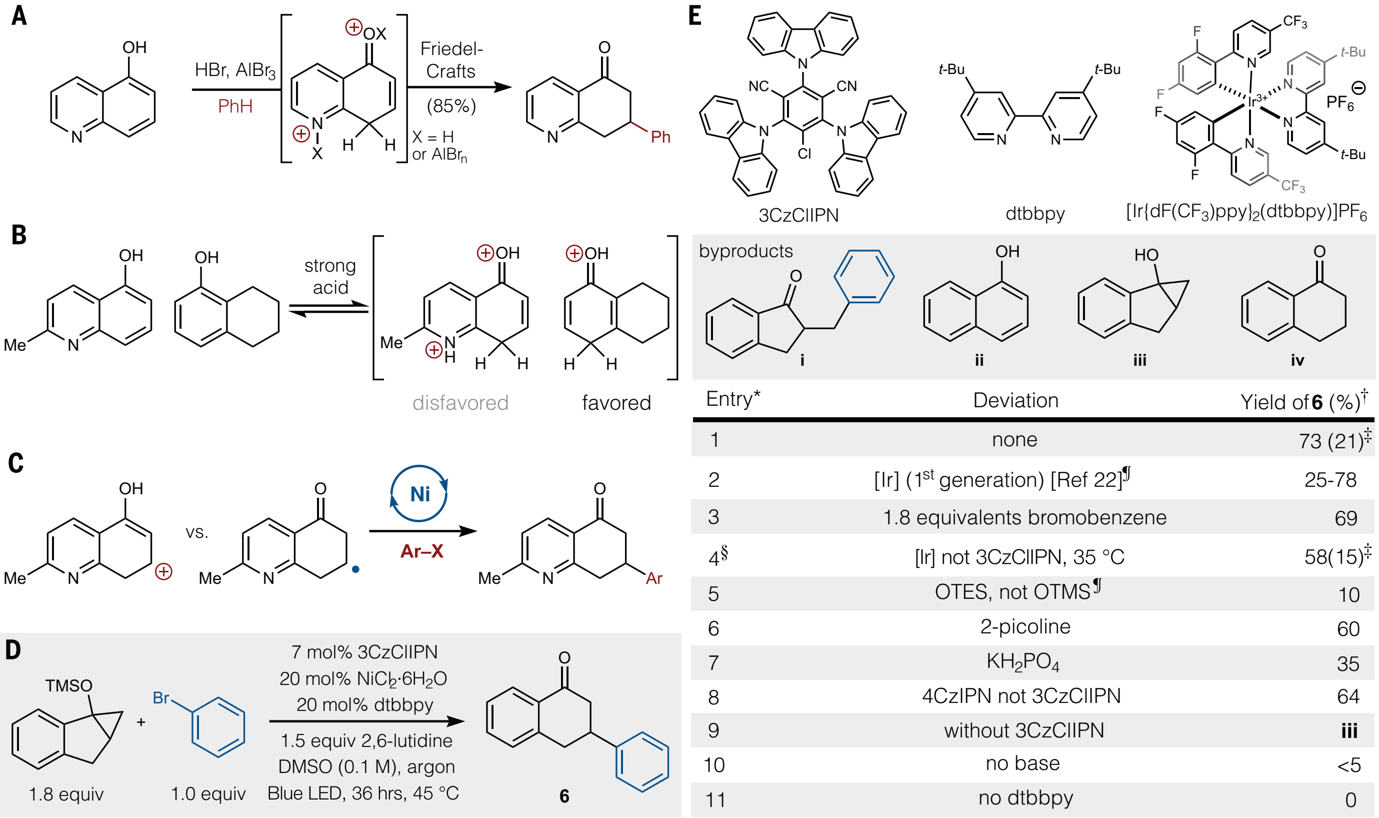


Fig. 1. Representative GB alkaloids and synthetic analysis. (**A**) Chemical space related to GB alkaloids. (**B**) Through-class retrosynthetic analysis relating GB alkaloids. (**C**) A high $F_{sp}2$, low $F_{sp}3$ intermediate that undergoes rapid retro–Friedel-Crafts attached-ring scission (supplementary materials) but can be formed under mild cross-coupling conditions (*20*).

(*26*). Yields of coupled product **6** depended on heat removal by means of air circulation and were difficult to reproduce without tight control of temperature (**ii** to **iv** in Fig. 2D were major byproducts). Ultimately, the reaction setup was altered to accommodate the use of a water bath for temperature control and heating. When organic photocatalysts such as 3CzClIPN and 4CzIPN were used in place of Ir, the reaction proved faster and more reliable, with lower costs and more flexibility over conditions (*27*). Highly polar solvents such as dimethylamine (DMA) and N,N′-dimethylformamide (DMF) were competent in this chemistry, but dimethyl sulfoxide (DMSO) outperformed both. Additionally, carbonate bases were ineffective, and phosphate bases proved inferior and less consistent than organic pyridine bases such as 2,6-lutidine and 2-picoline. The yield of **6** decreased when we used photocatalysts that had higher or lower oxidation potentials than 3CzClIPN. Last, the reaction did not proceed in controls that excluded each of the reagents (tables S1 to S4). These conditions were generally successful across a variety of siloxycyclopropanes and haloarenes (Fig. 3A). In all cases, retrosynthetic disconnections using a conjugate addition transform would lead to a cyclohexadienone that exists as the phenol tautomer (**ii**).

Neither electron-withdrawing nor electron-donating groups on the arene affected the efficiency of the coupling reaction, and heterocycles were also compatible [oxidant-sensitive arenes were problematic, however (supplementary materials, pages S20 to S31)]. The reaction translated from bromoarenes to 1-bromocyclohexene, albeit in reduced yield. Encouraged by the scope of this cross-coupling, especially with regard to heterocyclic substrates, we investigated entry into the synthesis of GB22 (Fig. 3B). Ketone **7** is available commercially and can be prepared in one step through condensation of 1,3-cyclopentanedione, methyl vinyl ketone, and ammonia (*28*). Conversion to siloxycyclopropane **8** was carried out by means of silyl enol ether formation, followed by Simmons-Smith cyclopropanation by using the Shi modification (*29*). An alternative three-step route to **8** from methyl 2-chloro-6-methyl-nicotinate was also developed to avoid the difficult purifications of the Simmons-Smith route. Cross-coupling with bromoarene **9a** or **9b** (two steps from 1-naphthol) (*30*) occurred cleanly, after optimization to account for *ortho*-substitution that leads to steric crowding of the intermediate arylnickel (tables S6 to S9). The higher yield of anisole **9b** likely reflects the low bond dissociation enthalpy of C–H bonds in benzyl ether **9a**.

The high fraction aromatic building block (high $F_{sp}2$) attached-ring intermediates **10a/b** incorporated all the skeletal carbons of GB22, GB13, and himgaline, but the projected Friedel-Crafts arylation proved difficult. First, as suggested by positive Hammett parameters (*31*), dominant inductive effects disfavored attack by the *meta*-position of the phenolic ether. Koltunov *et al.* found benzene itself to cyclize efficiently with 5-quinolol by using triflic acid (F_3CSO_3H) (*18*), but initial screens of strong Brønsted acids in our system delivered only small quantities of **4**. Triflic acid instead competitively protonated **10a/b** and effected a retro–Friedel-Crafts arylation to cleave the hard-won C–C bond and return the parent



Entry*	Deviation	Yield of **6** (%)†
1	none	73 (21)‡
2	[Ir] (1st generation) [Ref 22]¶	25-78
3	1.8 equivalents bromobenzene	69
4§	[Ir] not 3CzClIPN, 35 °C	58(15)‡
5	OTES, not OTMS¶	10
6	2-picoline	60
7	KH_2PO_4	35
8	4CzIPN not 3CzClIPN	64
9	without 3CzClIPN	**iii**
10	no base	<5
11	no dtbbpy	0

Fig. 2. Optimizing the arylation. (**A**) Failure of the Friedel-Crafts conjugate arylation. (**B**) Protonation of the wrong coupling partner. (**C**) Conceptual replacement of the cationic intermediate with a β-keto or -enol radical, leading to the development of the corresponding cross-coupling. (**D**) General conditions for attached-ring cross-coupling. (**E**) Optimization of catalysts and conditions. *0.1 mmol bromobenzene; ‡^{1}H-NMR yield; ‡(%yield of **i**); §0.25 mmol bromobenzene; and ¶with [Ir].

quinolol (Fig. 1B and table S5). Typical Lewis acids such as $AlCl_3$ also did not yield **4**. However, when inorganic aluminum Lewis acids were mixed with hexafluoroisopropanol (HFIP), tetracycle **4** was finally observed, albeit in low yield. We suspect that an aluminum species such as $Al[OCH(CF_3)_2]_nCl_m$ might act as an efficient Lewis acid (*32*) or hydrogen-bonding catalyst. Minimization of strong Brønsted acidity (HCl liberation) was accomplished by adding diethylaluminum chloride to HFIP, which quickly and exothermically evolved gas (likely ethane) to generate a new complex, tentatively assigned as $Al[OCH(CF_3)_2]_2Cl$ and its aggregates. The mechanism of cyclization may involve acidification of HFIP, formation of a strong double-hydrogen-bond donor bridged by aluminum, or formation of a strongly Lewis acidic complex (*33*). HFIP alone (*34*) did not promote any reaction of **10a**. This procedure led to clean and reproducible cyclization of the acid-labile attached-ring to either the parent phenol **4** (52%) or its methyl ether **11** (86%), depending on use of **10a** (to **4**) or **10b** (to **11**).

Both **4** and **11** could be hydrogenated over Rh/Al_2O_3 with exquisite stereocontrol to **12a/b** (other diastereomers not detected), analogously to related work on GB13 in which a larger, presaturated decalin motif provided steric shielding (*13*). Here, the benzene nucleus was unaffected by rhodium-catalyzed hydrogenation, but despite its planarity, small size, and capacity to adsorb to metal surfaces (*35*), it efficiently blocked the concave face of the pyridine ring. Whereas **12a** could be *N*-methylated [CH_2O (aqueous), $NaCNBH_3$] to GB22 directly, **12b** required demethylation by BBr_3 (75%), resulting in one more operation than that for the benzyl ether series but almost double the yield (16 versus 29%, three versus four steps). Proton nuclear magnetic resonance (^{1}H-NMR) and carbon-13 NMR (^{13}C-NMR) spectra of synthetic GB22 were identical to those reported by Lan and Mander (supplementary materials, p. S119) (*15*).

The next arene reduction benefited from retention of the methyl ether in **12b** and probed the role of the piperidine ring in control of stereochemistry at the incipient decalone. Birch reduction [Li^0 or Na^0 in NH_3 (liquid)] and electrochemical reduction proved unsuccessful, but Benkeser reduction (Li^0 in $MeNH_2$, THF/*i*-PrOH) effected clean reduction of the anisole (*36*). Proton source and metal proved crucial: MeOH, EtOH, and *t*-BuOH did not promote reduction, nor was Na^0 effective. The recently reported Koide reduction [Li^0, ethylenediamine, tetrahydrofuran (THF)] (*37*) worked extremely well and yielded similar amounts of product to that of $Li^0/MeNH_2$ with greater operational ease. A single diastereomer and regioisomer predominated (**14**) (x-ray crystallographic structure is provided in fig. S11), resulting from intermolecular protonation of C10 from the convex face, despite the potential for intramolecular proton transfer to C9 from the piperidine N–H, which was modeled to be only 2.46 Å apart in **12a** (x-ray). Minor byproduct pathways included demethylation (to **12a**), overreduction of the arene, and ~10% of a regioisomer. Hydrolysis of **14** with 2 M aqueous HCl, followed by basification (4 M NaOH), led to GB13 in 71% yield, with each of the remaining methine stereocenters adopting the desired relative configuration. Only a single methine (C10) positioned its hydrogen to the convex face, whereas two new stereocenters (C9 and C15) derived from prochiral, planar carbons that projected hydrogens inward. This stereochemistry may reflect, in part, the thermodynamic preferences of ring-tautomer (aza-Michael product) 16-oxohimgaline (**15**), which forms spontaneously under acidic conditions (*11*). Whereas the decalin *cis*-ring fusion of GB13 is calculated to be more stable by 1.3 kcal/mol, the *trans*-ring

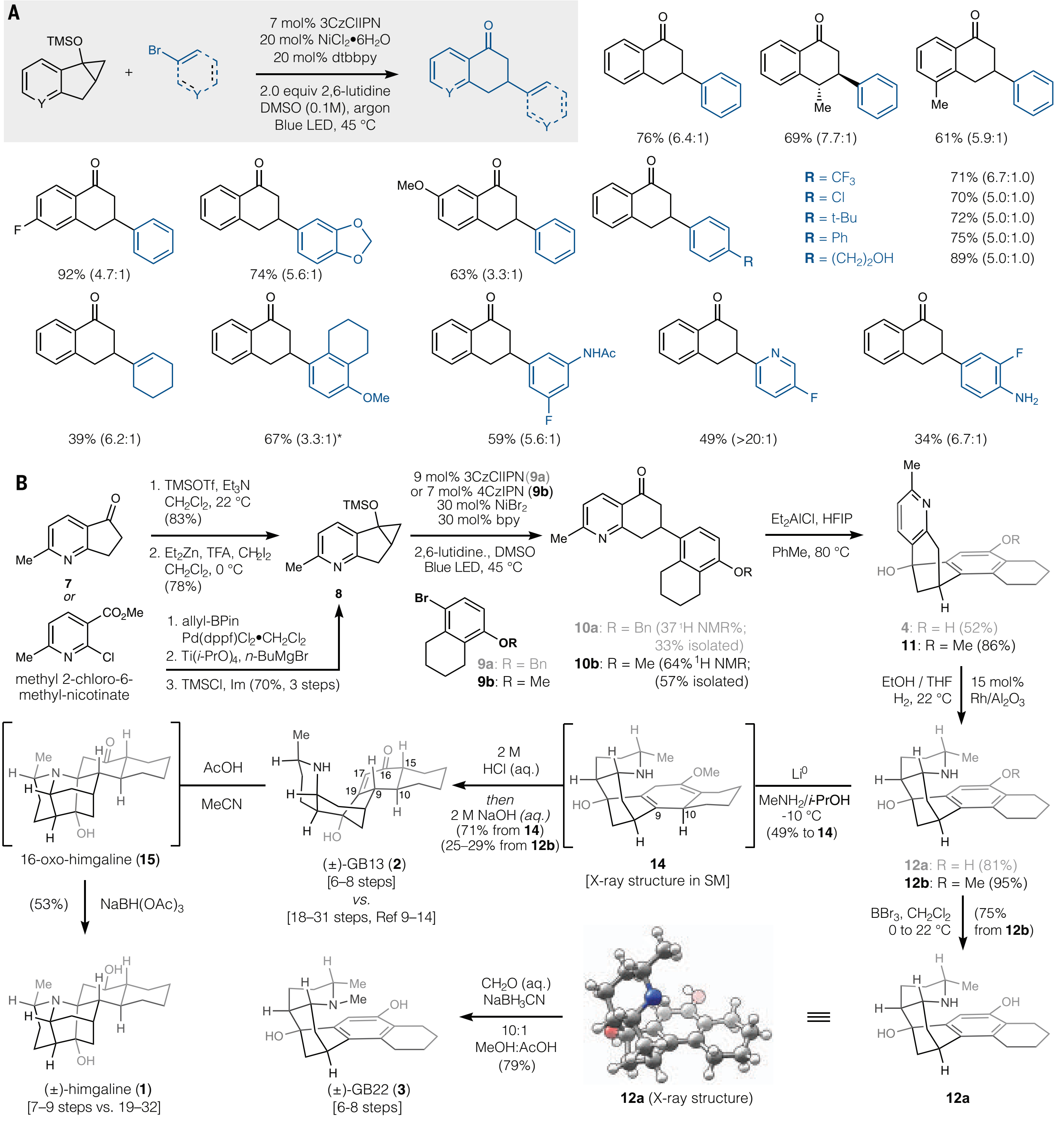


Fig. 3. Completion of the syntheses. (**A**) *Endo*-selective sp^3-sp^2 cross-coupling: percent isolated yield (*endo:exo*). *Same conditions as for **10a** (supplementary materials, pages S30 and S35). (**B**) A short synthesis of GB22, GB13, and himgaline through iterative, stereoselective reduction of multiple correlated carbons. 3CzClIPN, 2,4,6-tri(9H-carbazol-9-yl)-5-chloroisophthalonitrile; 4CzIPN, 1,2,3,5-tetrakis(carbazol-9-yl)-4,6-dicyanobenzene, 2,4,5,6-tetrakis(9H-carbazol-9-yl)isophthalonitrile; TFA, trifluoroacetic acid; TMSOTf, trimethylsilyl trifluoromethanesulfonate; bpy, 2,2′-bipyridine; HFIP, 1,1,1,3,3,3-hexafluoro-2-propanol; allyl-BPin, pinacolyl 2-propenylboronate; Pd(dppf)Cl_2•CH_2Cl_2, [1,1′-bis(diphenylphosphino)ferrocene]dichloropalladium(II), complex with dichloromethane; TMSCl, chlorotrimethylsilane; and Im, imidazole.

fusion of 16-oxo-himgaline is lower in energy by 2.7 kcal/mol (figs. S8 to S10). We speculate that the piperidine ammonium may deliver a proton internally to the enone γ-carbon C9 because β,γ to α,β-enone isomerization occurs under acidic conditions, and an extended enol tautomer is occluded on its concave face by this ammonium (fig. S12, **14**, x-ray). The final stereogenic methine C–H was installed according to a one-step protocol, as first demonstrated in a 33-step synthesis of himgaline (*12*). Thus, 10 prochiral atoms of high $F_{sp}2$ intermediate **10b** were converted in three steps to 10 new stereocenters (nine carbon and one nitrogen) by relay of increasing stereochemical information

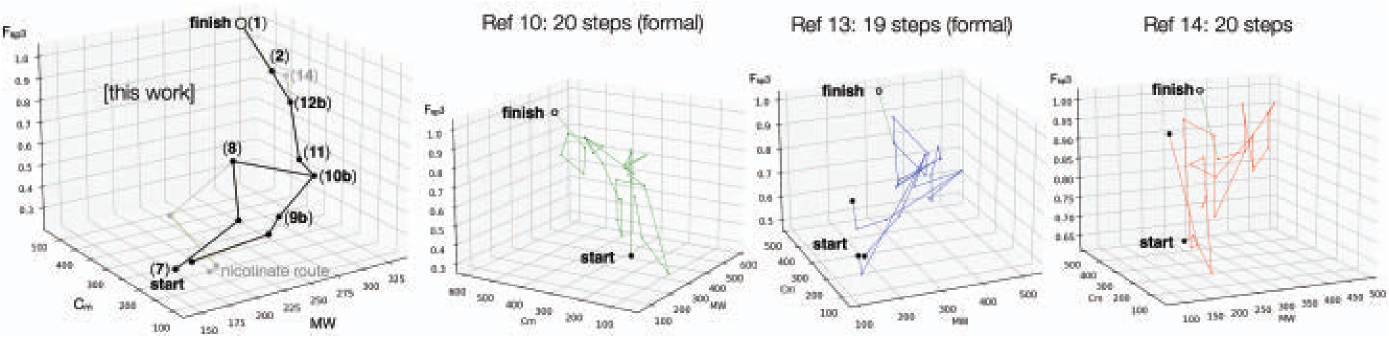


Fig. 4. Himgaline syntheses as walks through a chemical space. Plots are parameterized by $F_{sp}3$ (#$C_{sp}3/C_{total}$), C_m (mcbit), and molecular weight (daltons).

through simple reductions. To access pure GB13, **14** can be purified chromatographically to remove reduction by-products before acid hydrolysis, but this is unnecessary for conversion to pure himgaline, resulting in a seven- to nine-step synthesis, depending on isolation of **14** and designation of official starting material (**7** versus cyclopentane-1,3-dione versus methyl 2-chloro-6-methyl-nicotinate) (Fig. 4).

Himgaline is constitutionally related to cross-coupled product **10b** by these iterative additions of H_2, excluding the *O*-methyl embedded in starting material **9b**. Because hydrogen atoms are typically omitted from complexity calculations (*38*), the progression of high $F_{sp}2$ intermediates to 100% $F_{sp}3$ (himgaline) is exclusively due to information carried by molecular topology (C–C, C–N, and C–O bonds) and chirality content, properties captured with Böttcher's complexity measurement (C_m) of mcbits (*4*). Here, the 260.16 mcbits (*4*) of *des*-methyl **10b** increase to 477.83 over five steps, or 43.5 mcbits per step. Visualized as a walk through chemical space (Fig. 4) (*39*), the synthesis begins proximal to commercial space (low molecular weight, low complexity, and low $F_{sp}3$ /high $F_{sp}2$); converges early through cross-coupling; and then rapidly reaches the high complexity, weight, and $F_{sp}3$ of himgaline, which is typical of remote GB alkaloid space. By contrast, the shortest prior synthesis of himgaline (formal, racemic, and 19 steps) varies 148.45 mcbits over 15 steps (9.9 mcbits per step) from the latest point of convergency. Each route allows its own distinct exploration of different areas of chemical space. However, recognition that the key methine C–H stereocenters can be stereoselectively appended from prochiral sp^2 carbons of an aromatic himgaline core simplifies access to GB alkaloid space in a clear and quantifiable way. Whereas this analysis focuses on navigation to high-complexity chemotypes, we anticipate that GB structural chemical space can be better parameterized to relate to the biological targets and relative potencies among family members. With this goal in mind, we have scaled cross-coupling to 1.3 g of **10b**, the Friedel-Crafts to 1 g of **11**, and the hydrogenation to 834 mg of **12b** for diversification and target identification. Given the structural similarity among the 25 class II and III congeners, this approach is likely to prove general and to provide a means to deconvolute the targets, functions, and translational potential of the GB alkaloids.

REFERENCES AND NOTES

1. B. Thomas, *Sci. New Guinea* **25**, 33–59 (2000).
2. D. J. Collins, C. C. J. Culvenor, J. A. Lamberton, J. W. Loder, J. R. Price, in *Plants for Medicines* (Commonwealth Scientific and Industrial Research Organization, 1990), pp. 71–106.
3. U. Rinner, in *The Alkaloids: Chemistry and Biology* (Elsevier, 2017), vol. 78, chap. 2.
4. T. Böttcher, *J. Chem. Inf. Model.* **56**, 462–470 (2016).
5. S. A. Gilani, L. B. Cobbin, *Arch. Int. Pharmacodyn. Ther.* **290**, 46–53 (1987).
6. J. H. Miller, P. J. Aagaard, V. A. Gibson, M. McKinney, *J. Pharmacol. Exp. Ther.* **263**, 663–667 (1992).
7. S. Chackalamannil *et al.*, *J. Med. Chem.* **51**, 3061–3064 (2008).
8. S. Binns *et al.*, *Aust. J. Chem.* **18**, 569–573 (1965).
9. L. N. Mander, M. M. McLachlan, *J. Am. Chem. Soc.* **125**, 2400–2401 (2003).
10. M. Movassaghi, D. K. Hunt, M. Tjandra, *J. Am. Chem. Soc.* **128**, 8126–8127 (2006).
11. U. Shah *et al.*, *J. Am. Chem. Soc.* **128**, 12654–12655 (2006).
12. D. A. Evans, D. J. Adams, *J. Am. Chem. Soc.* **129**, 1048–1049 (2007).
13. K. K. Larson, R. Sarpong, *J. Am. Chem. Soc.* **131**, 13244–13245 (2009).
14. W. Zi, S. Yu, D. Ma, *Angew. Chem. Int. Ed.* **49**, 5887–5890 (2010).
15. P. Lan, A. J. Herlt, A. C. Willis, W. C. Taylor, L. N. Mander, *ACS Omega* **3**, 1912–1921 (2018).
16. E. R. Welin *et al.*, *Science* **363**, 270–275 (2019).
17. A. Ladenburg, *Ber. Dtsch. Chem. Ges.* **19**, 2578–2583 (1886).
18. K. Y. Koltunov, G. K. S. Prakash, G. Rasul, G. A. Olah, *J. Org. Chem.* **67**, 4330–4336 (2002).
19. S. A. Green, J. L. M. Matos, A. Yagi, R. A. Shenvi, *J. Am. Chem. Soc.* **138**, 12779–12782 (2016).
20. H. Rinderhagen, P. A. Waske, J. Mattay, *Tetrahedron* **62**, 6589–6593 (2006).
21. J. A. Milligan, J. P. Phelan, S. O. Badir, G. A. Molander, *Angew. Chem. Int. Ed.* **58**, 6152–6163 (2019).
22. H. E. Burdge, T. Oguma, T. Kawajiri, R. A. Shenvi, Concise synthesis of GB22 by *endo*-selective siloxycyclopropane arylation. chemRxiv 8263415.v1 [Preprint] (2019).
23. N. Varabyeva, M. Barysevich, Y. Aniskevich, A. Hurski, *Org. Lett.* **23**, 5452–5456 (2021).
24. D. D. Bume, C. D. Pitts, T. Lectka, *Eur. J. Org. Chem.* **2016**, 26–30 (2016).
25. S. Aoki, T. Fujimura, E. Nakamura, I. Kuwajima, *J. Am. Chem. Soc.* **110**, 3296–3298 (1988).
26. Z. Zuo *et al.*, *Science* **345**, 437–440 (2014).
27. E. Speckmeier, T. G. Fischer, K. Zeitler, *J. Am. Chem. Soc.* **140**, 15353–15365 (2018).
28. C. Allais, F. Liéby-Muller, J. Rodriguez, T. Constantieux, *Eur. J. Org. Chem.* **2013**, 4131–4145 (2013).
29. J. C. Lorenz *et al.*, *J. Org. Chem.* **69**, 327–334 (2004).
30. J. A. Lowe 3rd *et al.*, *J. Med. Chem.* **47**, 1575–1586 (2004).
31. C. Hansch, A. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology* (Wiley-Interscience, 1979).
32. K. S. Mazdiyasni, B. J. Schaper, L. M. Brown, *Inorg. Chem.* **10**, 889–892 (1971).
33. H. Yamamoto, K. Futatsugi, *Angew. Chem. Int. Ed.* **44**, 1924–1942 (2005).
34. H. F. Motiwala, R. H. Vekariya, J. Aubé, *Org. Lett.* **17**, 5484–5487 (2015).
35. S. Mitsui, S. Imaizumi, Y. Esashi, *Bull. Chem. Soc. Jpn.* **43**, 2143–2152 (1970).
36. R. A. Benkeser, C. Arnold Jr., R. F. Lambert, O. H. Thomas, *J. Am. Chem. Soc.* **77**, 6042–6045 (1955).
37. J. Burrows, S. Kamo, K. Koide, *Science* **374**, 741–746 (2021).
38. S. H. Bertz, *J. Am. Chem. Soc.* **103**, 3599–3601 (1981).
39. R. M. Demoret *et al.*, *J. Am. Chem. Soc.* **142**, 18599–18618 (2020).

ACKNOWLEDGMENTS

We thank L. Smith (Concordia College, Moorhead) for assistance with the synthesis of **8** and S. J. Hill for assistance with Matplotlib. We thank N. Jui for sending 3D-printed racks for photochemical reactions and C. Bi for help with high-pressure hydrogenation. We thank H. Renata and S. Crossley for proofreading. We thank A. Rheingold, M. Gembicky, and E. Samolova for x-ray crystallographic analysis. **Funding:** This work was supported by National Institutes of Health grant GM122606 (R.A.S.); National Science Foundation grant CHE1856747 (R.A.S.); Skaggs Graduate School fellowship (E.M.L.); Bristol Myers Squibb fellowship (M.A.B.); Jiangsu Industrial Technology Research Institute (JITRI) fellowship (M.A.B. and H.E.B.); Shionogi & Co. fellowship (T.O.); and Japan Society for the Promotion of Science (JSPS) fellowship 17J08551 (T.K.). **Author contributions:** R.A.S. directed the research. H.E.B. devised first-generation conditions for cross-coupling; E.M.L. and M.A.B. devised second-generation conditions. T.O. developed the synthesis of **3**; M.A.B., R.A.S., and E.M.L. developed conditions for reduction of **12b**; and E.M.L. completed, optimized, and scaled syntheses of **1** and **2**. E.M.L. and R.A.S. parameterized (*9–14*) as walks through a chemical space. R.A.S., E.M.L., and M.A.B. composed the manuscript and the supplementary materials. **Competing interests:** A provisional patent has been filed by E.M.L., M.A.B., T.O., and R.A.S.: "Process and Intermediates for Preparing GB13, GB22 and himgaline"; USPTO serial no. 63/293,295; filing date 23 December 2021; Scripps reference no. 2128.0 P/TSR2599P. **Data and materials availability:** Structural parameters from x-ray crystallography are available free of charge from the Cambridge Crystallographic Data Centre (**12a**, CCDC 2122295; **14**, 2123248). Experimental procedures, copies of NMR spectra, x-ray structure reports, coordinates, and Matplotlib Python code are available in the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn8343
Materials and Methods
Figs. S1 to S12
Tables S1 to S10
NMR Spectra
References (*40–43*)

24 December 2021; accepted 27 January 2022
10.1126/science.abn8343

URBAN EVOLUTION

Global urban environmental change drives adaptation in white clover

James S. Santangelo[1,2], Rob W. Ness[1,2], Beata Cohan[1], Connor R. Fitzpatrick[3], Simon G. Innes[4,1], Sophie Koch[1], Lindsay S. Miles[1,2], Samreen Munim[5,1], Pedro R. Peres-Neto[6], Cindy Prashad[1], Alex T. Tong[1], Windsor E. Aguirre[7], Philips O. Akinwole[8], Marina Alberti[9], Jackie Álvarez[10], Jill T. Anderson[11], Joseph J. Anderson[12], Yoshino Ando[13], Nigel R. Andrew[14], Fabio Angeoletto[15], Daniel N. Anstett[16], Julia Anstett[17,18], Felipe Aoki-Gonçalves[19], A. Z. Andis Arietta[20], Mary T. K. Arroyo[21,22], Emily J. Austen[23], Fernanda Baena-Díaz[24], Cory A. Barker[25], Howard A. Baylis[26], Julia M. Beliz[27,28], Alfonso Benitez-Mora[29], David Bickford[30], Gabriela Biedebach[30], Gwylim S. Blackburn[31], Mannfred M. A. Boehm[16], Stephen P. Bonser[32], Dries Bonte[33], Jesse R. Bragger[34], Cristina Branquinho[35], Kristien I. Brans[36], Jorge C. Bresciano[37], Peta D. Brom[38], Anna Bucharova[39], Briana Burt[40], James F. Cahill[41], Katelyn D. Campbell[25], Elizabeth J. Carlen[42], Diego Carmona[43], Maria Clara Castellanos[44], Giada Centenaro[45], Izan Chalen[10,46], Jaime A. Chaves[10,47], Mariana Chávez-Pesqueira[48], Xiao-Yong Chen[49,50], Angela M. Chilton[51], Kristina M. Chomiak[40], Diego F. Cisneros-Heredia[10,46], Ibrahim K. Cisse[40], Aimée T. Classen[52], Mattheau S. Comerford[53], Camila Cordoba Fradinger[54], Hannah Corney[55], Andrew J. Crawford[56], Kerri M. Crawford[57], Maxime Dahirel[58], Santiago David[59], Robert De Haan[60], Nicholas J. Deacon[61], Clare Dean[62], Ek del-Val[63], Eleftherios K. Deligiannis[64], Derek Denney[11], Margarete A. Dettlaff[41], Michelle F. DiLeo[65], Yuan-Yuan Ding[49], Moisés E. Domínguez-López[66,67], Davide M. Dominoni[68], Savannah L. Draud[69], Karen Dyson[9], Jacintha Ellers[70], Carlos I. Espinosa[71], Liliana Essi[72], Mohsen Falahati-Anbaran[73,74], Jéssica C. F. Falcão[75], Hayden T. Fargo[1], Mark D. E. Fellowes[76], Raina M. Fitzpatrick[77], Leah E. Flaherty[78], Pádraic J. Flood[79], María F. Flores[22], Juan Fornoni[80], Amy G. Foster[81], Christopher J. Frost[82], Tracy L. Fuentes[9], Justin R. Fulkerson[83], Edeline Gagnon[84,85], Frauke Garbsch[81], Colin J. Garroway[86], Aleeza C. Gerstein[87], Mischa M. Giasson[88], E. Binney Girdler[89], Spyros Gkelis[64], William Godsoe[90], Anneke M. Golemiec[5], Mireille Golemiec[1], César González-Lagos[29,91], Amanda J. Gorton[92], Kiyoko M. Gotanda[93,26], Gustaf Granath[12], Stephan Greiner[81], Joanna S. Griffiths[94], Filipa Grilo[35], Pedro E. Gundel[95,54], Benjamin Hamilton[40], Joyce M. Hardin[69], Tianhua He[96,97], Stephen B. Heard[88], André F. Henriques[35], Melissa Hernández-Poveda[56], Molly C. Hetherington-Rauth[1], Sarah J. Hill[14], Dieter F. Hochuli[98], Kathryn A. Hodgins[99], Glen R. Hood[100], Gareth R. Hopkins[101], Katherine A. Hovanes[102], Ava R. Howard[101], Sierra C. Hubbard[69], Carlos N. Ibarra-Cerdeña[103], Carlos Iñiguez-Armijos[71], Paola Jara-Arancio[104,105], Benjamin J. M. Jarrett[106,26], Manon Jeannot[107], Vania Jiménez-Lobato[108], Mae Johnson[109], Oscar Johnson[110], Philip P. Johnson[111], Reagan Johnson[112], Matthew P. Josephson[113], Meen Chel Jung[9], Michael G. Just[114], Aapo Kahilainen[65], Otto S. Kailing[115], Eunice Kariñho-Betancourt[116], Regina Karousou[64], Lauren A. Kirn[99], Anna Kirschbaum[117], Anna-Liisa Laine[118,65], Jalene M. LaMontagne[7,119], Christian Lampei[39], Carlos Lara[120], Erica L. Larson[121], Adrián Lázaro-Lobo[122], Jennifer H. Le[123], Deleon S. Leandro[124], Christopher Lee[99], Yunting Lei[125], Carolina A. León[29], Manuel E. Lequerica Tamara[98], Danica C. Levesque[126], Wan-Jin Liao[127], Megan Ljubotina[41], Hannah Locke[57], Martin T. Lockett[128], Tiffany C. Longo[34], Jeremy T. Lundholm[55], Thomas MacGillavry[68], Christopher R. Mackin[44], Alex R. Mahmoud[27], Isaac A. Manju[101], Janine Mariën[70], D. Nayeli Martínez[63,129], Marina Martínez-Bartolomé[130,122], Emily K. Meineke[131], Wendy Mendoza-Arroyo[116], Thomas J. S. Merritt[126], Lila Elizabeth L. Merritt[126], Giuditta Migiani[68], Emily S. Minor[111], Nora Mitchell[132,133], Mitra Mohammadi Bazargani[134], Angela T. Moles[32], Julia D. Monk[20], Christopher M. Moore[135], Paula A. Morales-Morales[136], Brook T. Moyers[137,138], Miriam Muñoz-Rojas[51,139], Jason Munshi-South[42], Shannon M. Murphy[121], Maureen M. Murúa[140], Melisa Neila[29], Ourania Nikolaidis[123], Iva Njunji[141], Peter Nosko[142], Juan Núñez-Farfán[80], Takayuki Ohgushi[143], Kenneth M. Olsen[27], Øystein H. Opedal[106], Cristina Ornelas[144], Amy L. Parachnowitsch[88,12], Aaron S. Paratore[40], Angela M. Parody-Merino[37], Juraj Paule[145], Octávio S. Paulo[35], João Carlos Pena[146], Vera W. Pfeiffer[147], Pedro Pinho[35], Anthony Piot[31], Ilga M. Porth[31], Nicholas Poulos[148], Adriana Puentes[149], Jiao Qu[33], Estela Quintero-Vallejo[150], Steve M. Raciti[151], Joost A. M. Raeymaekers[152], Krista M. Raveala[65], Diana J. Rennison[153], Milton C. Ribeiro[146], Jonathan L. Richardson[154], Gonzalo Rivas-Torres[10,155], Benjamin J. Rivera[89], Adam B. Roddy[156], Erika Rodriguez-Muñoz[56], José Raúl Román[157], Laura S. Rossi[142], Jennifer K. Rowntree[62], Travis J. Ryan[158], Santiago Salinas[89], Nathan J. Sanders[52], Luis Y. Santiago-Rosario[159], Amy M. Savage[123], J.F. Scheepens[160,117], Menno Schilthuizen[161], Adam C. Schneider[69,1], Tiffany Scholier[149,162], Jared L. Scott[163], Summer A. Shaheed[34], Richard P. Shefferson[164], Caralee A. Shepard[69], Jacqui A. Shykoff[165], Georgianna Silveira[166], Alexis D. Smith[111], Lizet Solis-Gabriel[63], Antonella Soro[167], Katie V. Spellman[168,144], Kaitlin Stack Whitney[169], Indra Starke-Ottich[145], Jörg G. Stephan[170,149], Jessica D. Stephens[171], Justyna Szulc[172], Marta Szulkin[172], Ayco J. M. Tack[45], Ítalo Tamburrino[22], Tayler D. Tate[101], Emmanuel Tergemina[79], Panagiotis Theodorou[167], Ken A. Thompson[59,173], Caragh G. Threlfall[98], Robin M. Tinghitella[121], Lilibeth Toledo-Chelala[63], Xin Tong[49], Léa Uroy[58,174], Shunsuke Utsumi[13], Martijn L. Vandegehuchte[107,33], Acer VanWallendael[175], Paula M. Vidal[22], Susana M. Wadgymar[176], Ai-Ying Wang[127], Nian Wang[177], Montana L. Warbrick[142], Kenneth D. Whitney[132], Miriam Wiesmeier[178], J. Tristian Wiles[69], Jianqiang Wu[125], Zoe A. Xirocostas[32], Zhaogui Yan[177], Jiahe Yao[179], Jeremy B. Yoder[148], Owen Yoshida[55], Jingxiong Zhang[125], Zhigang Zhao[179], Carly D. Ziter[6], Matthew P. Zuellig[180], Rebecca A. Zufall[57], Juan E. Zurita[10], Sharon E. Zytynska[178,181], Marc T. J. Johnson[1,2]*†

Urbanization transforms environments in ways that alter biological evolution. We examined whether urban environmental change drives parallel evolution by sampling 110,019 white clover plants from 6169 populations in 160 cities globally. Plants were assayed for a Mendelian antiherbivore defense that also affects tolerance to abiotic stressors. Urban-rural gradients were associated with the evolution of clines in defense in 47% of cities throughout the world. Variation in the strength of clines was explained by environmental changes in drought stress and vegetation cover that varied among cities. Sequencing 2074 genomes from 26 cities revealed that the evolution of urban-rural clines was best explained by adaptive evolution, but the degree of parallel adaptation varied among cities. Our results demonstrate that urbanization leads to adaptation at a global scale.

Urbanization is a driver of both environmental and evolutionary change. Towns and cities are rapidly expanding throughout the world to accommodate human population growth. These urban areas represent novel ecosystems, in which urban development alters multiple environmental factors (*1*). Recent research shows that urban environmental change can influence four evolutionary processes: mutation, genetic drift, gene flow, and adaptation due to natural selection (*2*, *3*). Despite numerous examples of how urbanization affects genetic drift and gene flow (*4*, *5*), the effects of urbanization on adaptive evolution have received less attention (*6*–*8*). Adaptation to urban environments can affect species' conservation (*9*), the spread of pests and disease (*2*), and eco-evolutionary feedbacks (*10*), as well as urban planning and human society (*11*). However, the few examples of adaptation to urban environments focus on just one or a small number of cities in a single region (*2*). It is therefore unclear whether populations can adapt to urban habitats in similar ways across cities throughout the world.

Parallel adaptive evolution is most likely when populations experience similar environmental selective pressures on the same genes or phenotypes (*12*, *13*). For urbanization to drive parallel evolution, urban areas must converge in environmental features that affect

---

*Corresponding author. Email: marc.johnson@utoronto.ca
†Affiliations are listed at the end of this paper.

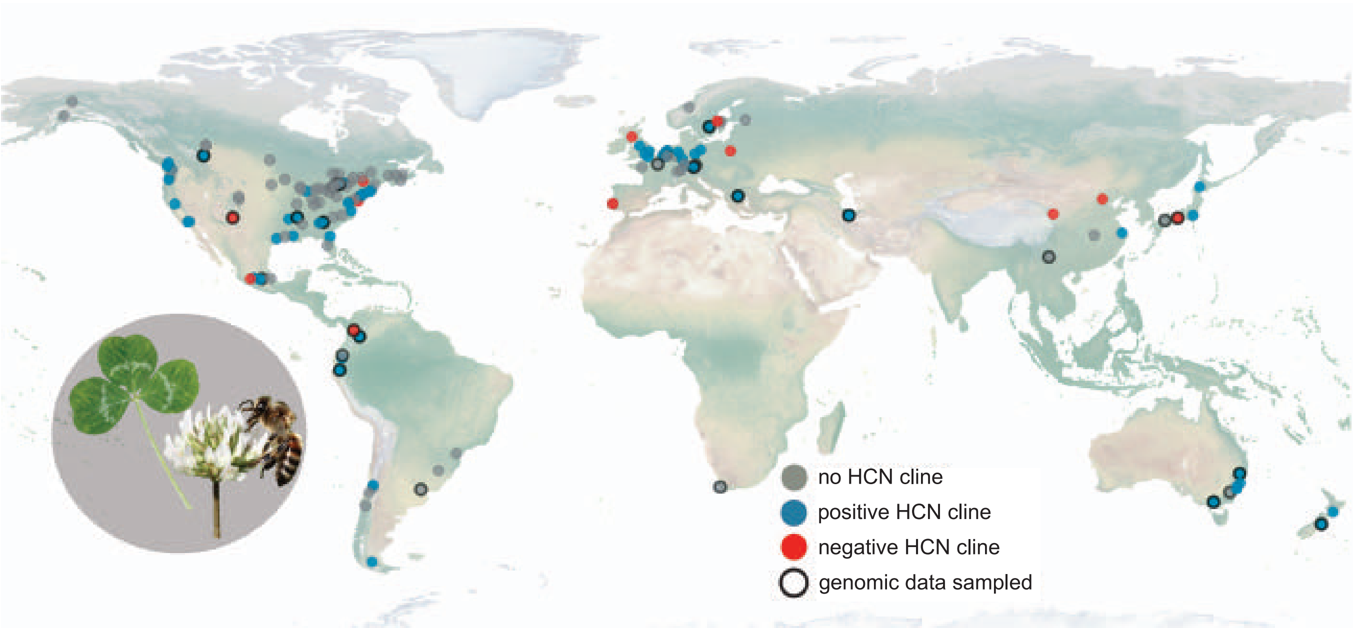


Fig. 1. Cities sampled for urban environmental and evolutionary change. Blue dots indicate cities with positive clines for hydrogen cyanide (HCN) production along urban-rural gradients ($HCN_{urban} < HCN_{rural}$). Red dots show negative clines ($HCN_{urban} > HCN_{rural}$). Gray dots indicate cities without a cline. Plants from the 26 cities outlined in black underwent whole-genome sequencing. Inset: White clover and a honey bee.

an organism's fitness. Urbanization can lead to similar environmental changes across cities (*14*), but whether urban environmental convergence causes parallel evolution has never been examined at a global scale.

Here we test how global urbanization affects environmental change and evolution in a cosmopolitan plant species, white clover (*Trifolium repens* L., Fabaceae). White clover populations are polymorphic for the production of hydrogen cyanide (HCN), an antiherbivore chemical defense controlled by two genes (*15*). At least one functional allele at each of two unlinked loci (*Ac* and *Li*) is required to produce HCN following tissue damage, whereas plants that are homozygous for gene deletions (*ac* and *li* alleles) at either locus lack HCN (*16*, *17*). Notably, these deletions occur throughout the world, resulting in standing genetic variation on which selection can act (*18*). Previous work showed that herbivores select for the production of HCN, and abiotic stressors (e.g., freezing and drought) influence the costs and benefits of the metabolic components underlying the defense (*19*, *20*). Variation in these environmental factors is credited with driving the evolution of clines in HCN production at continental and regional scales (*21*, *22*), including in response to urban environments (*23*–*25*). Thus, HCN production could evolve in response to urbanization if there are urban-rural gradients in herbivory, winter temperature, or drought.

We examined global urban environmental and evolutionary change across the diverse climates that white clover inhabits. To this end, we created the Global Urban Evolution Project to test for parallel evolution and urban adaptation in natural populations across white clover's worldwide range. The present study builds on our previous work on white clover (*23*–*25*) by sampling cities globally across diverse climates in both the native (Europe and western Asia) and introduced ranges, by quantifying many environmental factors from each population and by integrating evolutionary genomic analyses using whole-genome sequence data. This project spanned 160 cities across 26 countries (Fig. 1) (*15*) in white clover's native and introduced ranges (Fig. 1 and fig. S1). From these cities, we phenotyped 110,019 plants from 6169 sampling sites (hereafter "populations," table S1). Populations within each city were sampled along an urban-rural transect, with half of each transect in urban and suburban areas (i.e., areas with high building density) and the other half in rural areas (Fig. 2, E to G) (*15*).

Across 160 cities, we tested whether urban white clover habitats converged to be more similar among cities and less variable within cities in their environmental characteristics compared to rural habitats (*15*). Urban and rural habitats significantly diverged (MANOVA H_0: $urban_{mean} = rural_{mean}$, $P_{bootstrapped} < 0.01$, Fig. 2A) along two principal-component axes that accounted for 65% of the variation in the multivariate environments between the two habitats across cities. Urban locations consistently had more impervious surface, higher summer temperatures, and less vegetation than rural populations (Fig. 2B and fig. S2). The remaining environmental variables changed along urban-rural gradients in many cities, but these changes were less consistent in direction among cities (fig. S2 and table S2). Although urban and rural environments diverged on average, urban-rural changes in the environment were not always parallel (MANOVA H_0: parallel urban-rural changes among cities, $P_{bootstrapped} < 0.01$, Fig. 2A). Additionally, environmental variance among urban populations within a city was lower than the environmental variance among rural populations ($F_{9,1570} = 31.76$, $P < 0.001$, fig. S3). Together these results show that on average, urbanization leads to similar and less-variable environmental conditions in some factors (e.g., impervious surface, summer temperature, summer vegetation) but not in others (e.g., potential evapotranspiration, snow cover, winter vegetation), which could lead to variation in the degree of parallel evolution.

We next tested whether convergent urban environmental change causes parallel evolution in an ecologically important trait of white clover. We examined evolution in response to urbanization by testing for a relationship between HCN production and distance to the urban center (i.e., an "HCN cline"), as well as other metrics of urbanization (*15*). Our model explained 28% of the variation in the frequency of HCN production within populations (table S3). Across 160 cities, distance from the city center was positively related to the frequency of HCN-producing plants (distance: $\chi^2_{df=1} = 12.35$, $P < 0.001$). The probability that a plant produced HCN increased by 44% on average from the center of an urban area to the furthest rural population (Fig. 2C, D). However, cities varied in the strength and direction

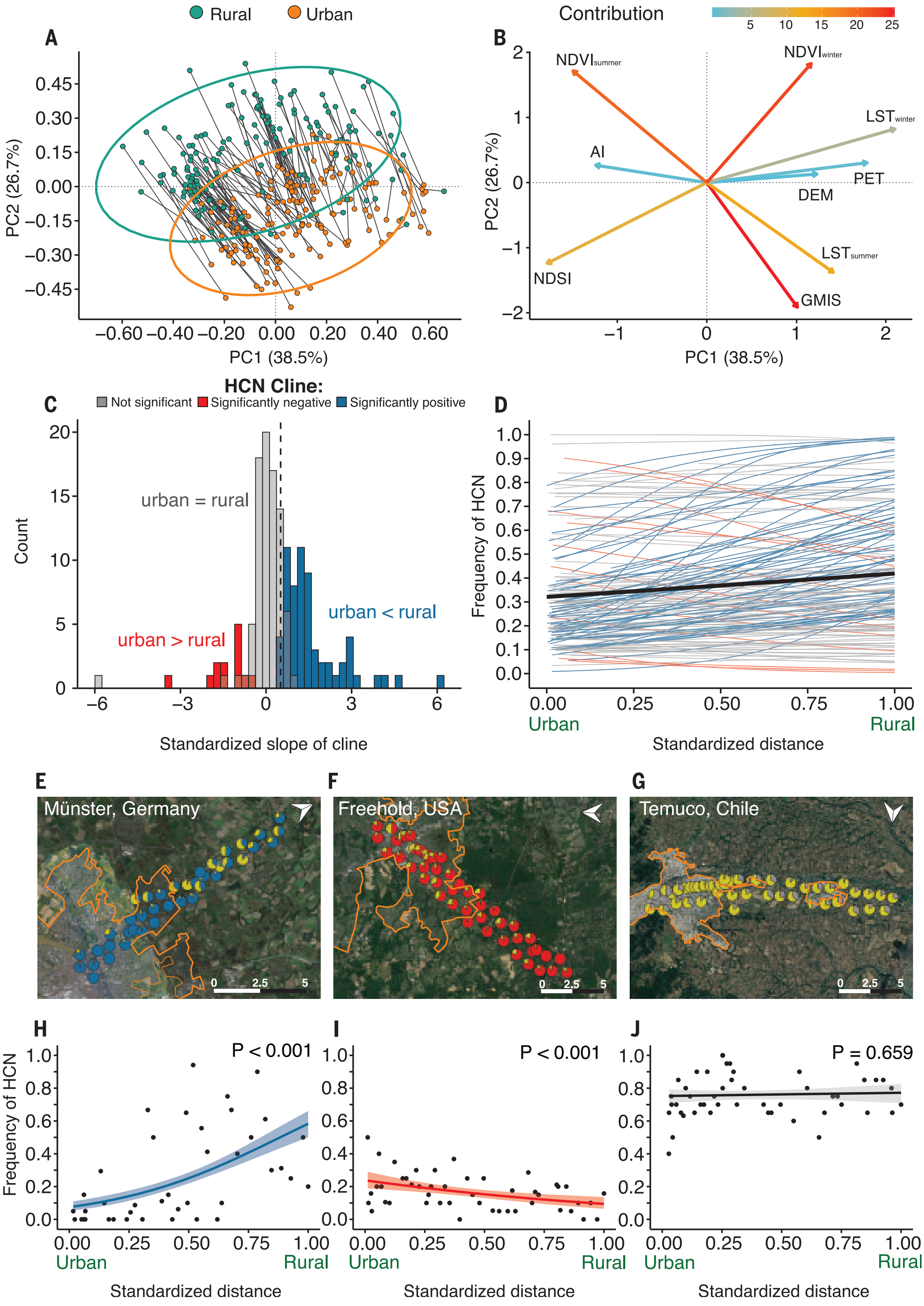


Fig. 2. Urban environmental and evolutionary change across cities. (**A**) Principal component analysis showing environmental differences between urban (orange dots) and rural (green dots) habitats; ovals represent 95% confidence interval (CI). Lines connect urban and rural habitats from the same city. (**B**) The eigenvectors for environmental variables, colored according to their contribution to PC2. The environmental variables included vegetation in winter ($\mathrm{NDVI}_{\mathrm{winter}}$) and summer ($\mathrm{NDVI}_{\mathrm{summer}}$), snow accumulation (NDSI), surface temperature in winter ($\mathrm{LST}_{\mathrm{winter}}$) and summer ($\mathrm{LST}_{\mathrm{summer}}$), aridity index (AI), potential evapotranspiration (PET), impervious surface (GMIS), and elevation (DEM). (**C**) Histogram of the slopes from binomial regressions of the relationship between HCN production within populations and distance from the city center. Distance was standardized to vary between 0 (urban center) and 1 (furthest rural population) in each city, so that cities that varied in size were compared on the same scale. The dashed vertical line corresponds to the mean slope across cities, and overlap between bars showing cities with significant (blue and red) and nonsignificant clines (grey) is shown as muted colors. (**D**) The relationship between HCN production within populations and distance for each city; colors correspond to those in (C). The black line shows the positive main effect of distance across cities ($P < 0.001$). (**E** to **G**) Examples of transects, with the orange lines showing the urban boundary, and pie charts (jittered to reduce overlap) showing the proportion of HCN+ plants colored in yellow. (**H** to **J**) Frequency of HCN production versus distance for the cities shown in (E) to (G). The line shows the regression line ± 95% CI.

of clines (distance × city interaction: $\chi^2_{\mathrm{df}=1}$= 1001, $P < 0.001$; Fig. 2, C and D). Overall, 47% of cities exhibited a significant ($P < 0.05$) cline (*15*), with 39% of cities (62 of 160) showing a positive cline in which HCN production was less common in urban than rural populations, and 8% of cities (13 of 160) had negative clines (Fig. 2 and table S4). Positive and negative clines occurred in both the native and introduced ranges, with the former being more prevalent among continents and across diverse climates (Fig. 1).

Given the prevalence of HCN clines at a global scale, we sought to identify the evolutionary processes driving variation in the strength and direction of clines. In addition to natural selection, nonadaptive evolution can lead to the evolution of clines (*26*). Notably, the epistatic genetic architecture of HCN production makes the loss of the trait more likely with increased genetic drift (*26*). Therefore, the prevalence of positive clines could reflect stronger drift in urban populations (*4*, *5*). To examine whether urban populations exhibited stronger drift, we estimated pairwise nucleotide diversity (π) of putatively neutral sites using whole-genome sequence data from ~80 individuals per city, with samples equally split between urban and rural habitats across 26 cities (*N* = 2,074) (*15*). These cities were selected to capture variation in the strength and direction of clines, geography, and climate (Fig. 1) (*15*).

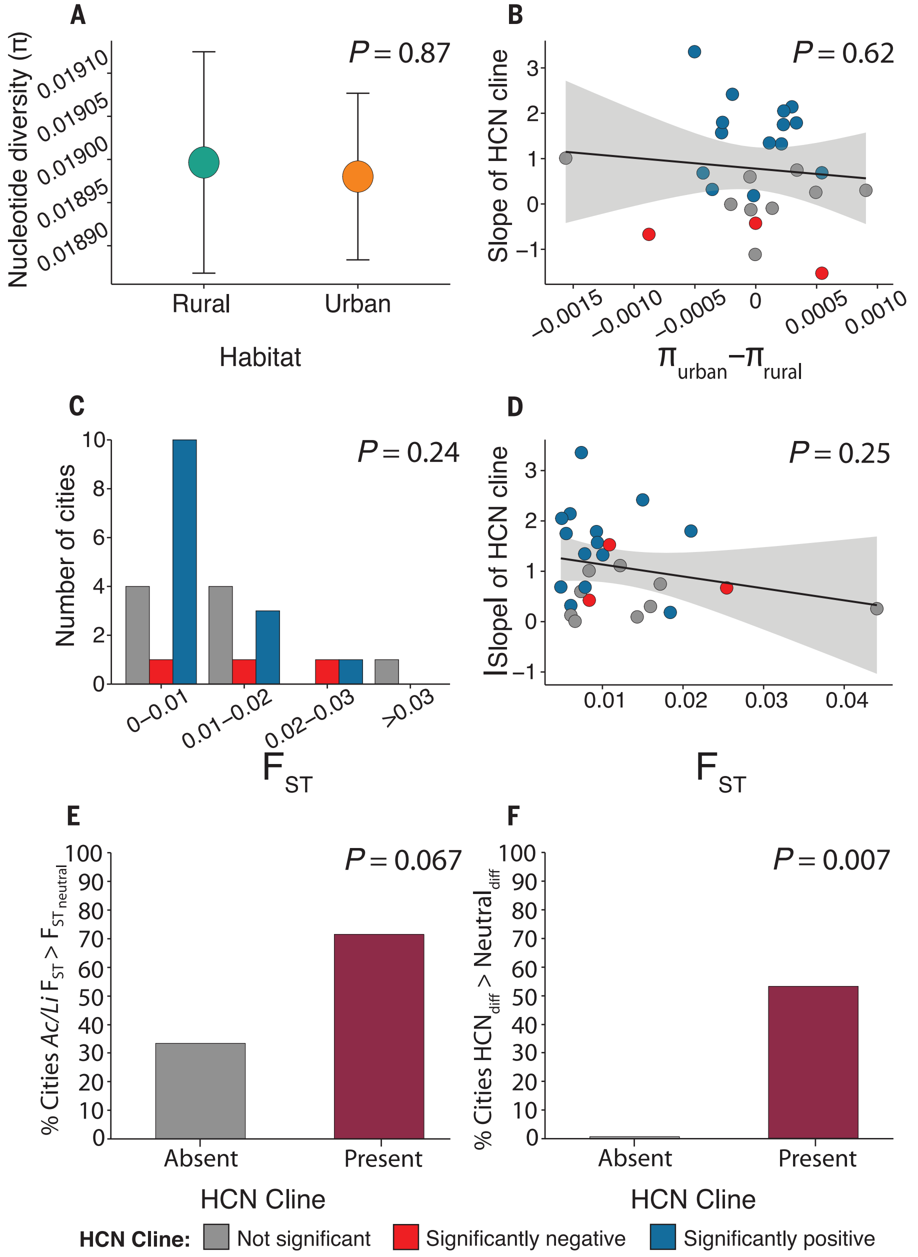


Fig. 3. Genetic diversity and differentiation within and between urban and rural habitats. (**A**) Mean (± SE) pairwise nucleotide diversity (π) for urban (orange) and rural (green) plants across cities. (**B**) The relationship between the slope of HCN clines versus the difference in nucleotide diversity between habitats, where each point is a city. (**C**) Histogram showing the distribution of genetic differentiation (F_{ST}) between urban and rural habitats for each city, colored with respect to the significance of HCN clines. (**D**) Relationship between the absolute value of the slope of HCN clines versus F_{ST}. (**E**) Percentage of cities in which differentiation between urban and rural habitats at *Ac* or *Li* exceeds neutral expectation in cities with or without significant HCN clines (*15*). (**F**) Percentage of cities with differentiation in HCN production between urban and rural habitats that exceeds neutral expectation in cities with or without significant HCN clines (*15*). *P* values in (E) and (F) correspond to χ^2 test for independence.

Genetic diversity was not consistently different between urban and rural habitats and did not explain variation in the slope of HCN clines along urban-rural gradients. On average, urban and rural habitats did not differ in neutral genetic diversity ($F_{1,\ 25}$ = 0.028, *P* = 0.87; Fig. 3A). Furthermore, the difference in π between urban and rural habitats within a city was not strongly related to the slope of HCN clines ($F_{1,\ 24}$ = 0.25, *P* = 0.62; Fig. 3B and fig. S4), and urban-rural differences in genetic diversity were similar between cities with and without clines ($F_{1,\ 24}$ = 0.017, *P* = 0.90).

Variation in the strength of genetic differentiation and gene flow between urban and rural habitats can influence the ability of populations to adapt to urban environments (*27*). To test the association between genetic differentiation and the evolution of HCN clines, we estimated population genetic differentiation between urban and rural populations using both F_{ST} and principal components analysis (PCA) (fig. S5), in addition to urban-rural admixture (fig. S6) (*15*). Urban-rural F_{ST} was low [mean = 0.012 ± 0.002 (SE)] and did not differ significantly between cities with and without clines ($F_{1,\ 24}$ = 1.47, *P* = 0.24; Fig. 3C and fig. S4). Neither F_{ST} ($F_{1,\ 24}$ = 1.42, *P* = 0.25; Fig. 3D) nor urban-rural differentiation measured using PCA ($F_{1,\ 24}$ = 1.10, *P* = 0.31, fig. S5) predicted the strength of clines in HCN production. The absence of strong differentiation was associated with extensive admixture between urban and rural populations (fig. S6). Because genetic differentiation is consistently low and gene flow appears to be high among urban and rural populations, the repeated evolution of clines suggests strong selection on HCN production along urban-rural gradients. This conclusion is further supported by direct tests of selection on the *Ac* and *Li* loci, as well as HCN production, in which differentiation (using a statistic equivalent to F_{ST}) between urban and rural populations was stronger than expected under neutral evolution in cities with HCN clines compared to cities without clines (Fig. 3, E and F) (*15*).

Multiple environmental stressors are known to influence the evolution of HCN production at continental scales (*20*–*22*, *28*), so we asked: What environmental factors explain variation in the evolution of HCN production along urban-rural gradients? Environmental factors related to drought and vegetation cover were the strongest predictors of variation in HCN clines, accounting for 11.3% of the variation in the strength of clines (tables S5 and S6). Change in potential evapotranspiration (PET) along urban-rural gradients was one of the

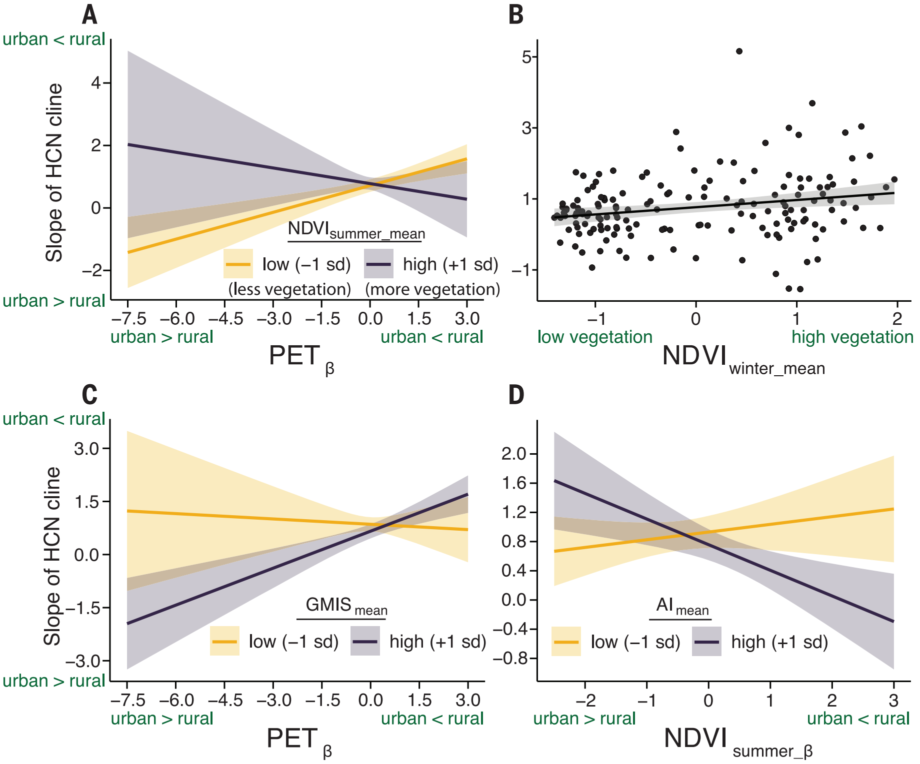


Fig. 4. Environmental predictors of urban-rural clines in HCN production. (**A**) Change in potential evapotranspiration along urban-rural gradients (PET$_\beta$) interacts with the regional amount of summer vegetation (i.e., NDVI$_{summer\_mean}$) to explain variation in the slopes of HCN clines. (**B**) The relationship between the slopes of HCN clines and the regional amount of winter vegetation (NDVI$_{winter\_mean}$). (**C**) PET$_\beta$ interacts with the regional amounts of impervious surface (GMIS$_{mean}$) to predict the slope of HCN clines. (**D**) Change in summer vegetation along urban-rural gradients (NDVI$_{summer\_\beta}$) interacts with regional aridity (AI$_{mean}$) to explain variation in the slope of HCN clines. Acronyms as in Fig. 2.

most consistent predictors of evolution in HCN production (table S5); the frequency of HCN production tended to be higher in rural than urban populations in cities where PET was also greater in rural habitats (Fig. 4, A and C, and fig. S7). Because high PET can lead to plant water stress under low soil moisture, this result is consistent with drought selecting for higher HCN production, a pattern also observed at continental scales (*21*). However, the effect of PET on the evolution of HCN production only occurs when the amount of vegetation in and around cities is low (Fig. 4A). When vegetation cover is relatively high (and impervious surface is low) along the whole urban-rural transect, HCN clines tend to be positive regardless of variation in PET (Fig. 4, A to C). Notably, the amount of vegetation is positively correlated with invertebrate herbivore biomass and diversity (*29*), which can select for increased HCN production (*20*). When combined with the observation that herbivores are often less abundant in urban habitats (*30*), our evidence suggests that herbivores are selecting for greater HCN production in rural than urban areas. The positive association between urban-rural changes in vegetation and the positive slope of HCN clines in some cities further supports this interpretation (Fig. 4D). Put simply, herbivory seems to select for higher HCN production in rural areas, but in the absence of strong herbivore pressure (i.e., when there is less vegetation across the whole gradient), drought is the main selective agent. Contrary to previous findings, urban-rural changes in temperature and snow cover did not explain changes in HCN production (*24*), suggesting that urban-rural changes in these abiotic factors are not a general explanation for the evolution of clines at a global scale.

Our results have general implications for understanding how environmental change affects adaptation in widespread species. Parallel evolution is a hallmark of natural selection because it suggests that adaptation proceeds in a repeatable way when populations face similar environments (*12*, *13*). However, departures from parallel evolution are common, and a major goal of recent research involves quantifying how ecological and evolutionary factors interact to influence variation in adaptive responses to similar environments (*12*). Our results show that white clover rapidly adapts to urban environments on a global scale, but there is considerable variation in the strength and direction of HCN clines that is driven by variation in particular biotic and abiotic factors that differ in how they change along urban-rural gradients among cities. Variation in additional unmeasured factors (e.g., gene flow from agricultural varieties, pollution, etc.) might further explain variation in the strength of clines, and future work will seek to explore such mechanisms.

Urbanization is increasingly transforming rural and natural environments into unique ecosystems that Earth's biodiversity has never experienced, and these changes are altering the evolution of life. If adaptation to urban environments is common, then this could have cascading effects on populations and ecosystems. This knowledge could help conserve some of Earth's most vulnerable species (*9*), mitigate the impacts of pests (*2*), improve human well-being (*8*, *11*), and contribute to understanding fundamental eco-evolutionary processes (*10*).

REFERENCES AND NOTES

1. N. B. Grimm *et al.*, *Science* **319**, 756–760 (2008).
2. M. T. J. Johnson, J. Munshi-South, *Science* **358**, eaam8327 (2017).
3. M. Szulkin, J. Munshi-South, A. Charmantier, Eds., *Urban Evolutionary Biology* (Oxford University Press, 2020).
4. L. S. Miles, L. R. Rivkin, M. T. J. Johnson, J. Munshi-South, B. C. Verrelli, *Mol. Ecol.* **28**, 4138–4151 (2019).
5. C. Schmidt, M. Domaratzki, R. P. Kinnunen, J. Bowman, C. J. Garroway, *Proc. Biol. Sci.* **287**, 20192497 (2020).
6. E. M. Oziolor *et al.*, *Science* **364**, 455–457 (2019).
7. K. M. Winchell, R. G. Reynolds, S. R. Prado-Irwin, A. R. Puente-Rolón, L. J. Revell, *Evolution* **70**, 1009–1022 (2016).
8. L. R. Rivkin *et al.*, *Evol. Appl.* **12**, 384–398 (2019).
9. M. R. Lambert, C. M. Donihue, *Nat. Ecol. Evol.* **4**, 903–910 (2020).
10. M. Alberti, *Trends Ecol. Evol.* **30**, 114–126 (2015).
11. C. J. Schell *et al.*, *Science* **369**, eaay4497 (2020).
12. D. I. Bolnick, R. D. Barrett, K. B. Oke, D. J. Rennison, Y. E. Stuart, *Annu. Rev. Ecol. Evol. Syst.* **49**, 303–330 (2018).
13. J. B. Losos, *Evolution* **65**, 1827–1840 (2011).
14. P. M. Groffman *et al.*, *Front. Ecol. Environ.* **12**, 74–81 (2014).
15. See supplementary materials.
16. K. M. Olsen, L. L. Small, *New Phytol.* **219**, 757–766 (2018).
17. K. M. Olsen, B. L. Sutherland, L. L. Small, *Mol. Ecol.* **16**, 4180–4193 (2007).
18. N. J. Kooyers, K. M. Olsen, *J. Evol. Biol.* **27**, 2554–2558 (2014).
19. N. J. Kooyers, B. Hartman Bakken, M. C. Ungerer, K. M. Olsen, *Am. J. Bot.* **105**, 1224–1231 (2018).
20. M. Hughes, *Heredity* **66**, 105–115 (1991).
21. N. J. Kooyers, L. R. Gage, A. Al-Lozi, K. M. Olsen, *Mol. Ecol.* **23**, 1053–1070 (2014).
22. H. Daday, *Heredity* **12**, 169–184 (1958).
23. M. T. J. Johnson, C. M. Prashad, M. Lavoignat, H. S. Saini, *Proc. Biol. Sci.* **285**, 20181019 (2018).
24. K. A. Thompson, M. Renaudin, M. T. J. Johnson, *Proc. Biol. Sci.* **283**, 20162180 (2016).
25. J. S. Santangelo *et al.*, *Evol. Lett.* **4**, 212–225 (2020).
26. J. S. Santangelo, M. T. J. Johnson, R. W. Ness, *Proc. Biol. Sci.* **285**, 20180230 (2018).
27. T. Lenormand, *Trends Ecol. Evol.* **17**, 183–189 (2002).
28. N. J. Kooyers, K. M. Olsen, *Heredity* **111**, 495–504 (2013).
29. M. Fernández-Tizón, T. Emmenegger, J. Perner, S. Hahn, *Naturwissenschaften* **107**, 42 (2020).
30. L. S. Miles, S. T. Breitbart, H. H. Wagner, M. T. J. Johnson, *Front. Ecol. Evol.* **7**, 310 (2019).
31. J. S. Santangelo, James-S-Santangelo/glue_pc: Minor documentation and model updates, version 1.0.1, Zenodo (2021); https://doi.org/10.5281/zenodo.5780438.

ACKNOWLEDGMENTS

We thank L. Alejandro Giraldo, L. Arboleda-Restrepo, E. Bernal, F. Carrera, M. T. Solano de la Cruz, K. Christensen-Dalsgaard, K. Cuypers, E. Dawaas, A. Giraldo, D. González-Tokman, B. Gravendeel, T. Gregor, J. Hatakoshi, P. Hyttinen, S. Kagiya, H. Kappes, B. Kerr, A. Matsuura, S. Silberhorn, B. Kwan, M. Potter, E. Peñaherrera, J. Rafalski, L. Revell, E. Sparrow, R. Tapia-López, A. Tovar, Y. Wang, J. Wrath, L. Yaneva-Roder, X. Zhu, 2018 MacEwan University BIOL422 students, Minneapolis College's 2018 Plant Biology students, SHAD Mount Allison 2018 students and staff, and University of Wisconsin-Madison's 2018 Field Ecology students for assisting with collecting plants, performing HCN assays, or providing equipment and facilities. D. Murray-Stoker, A. Filazzola, L. Albano, S. Breitbart, and R. Rivkin provided comments on an earlier draft of the paper. I. Sheoran prepared most genomic libraries. M. Malcolm and X. Xhao assisted with shipping and lab logistics, respectively. High-performance computing services were provided by Compute Ontario (www.computeontario.ca/) and Compute Canada (www.computecanada.ca). This work benefited from ideas and collaborations in Future Earth's EvolvES network and the NSF-funded RCN Urban Eco-Evo NET. **Funding:** The Global Urban Evolution project was primarily funded by an NSERC Discovery Grant, Canada Research Chair and NSERC Steacie Fellowship to M.T.J.J.. J.S.S. received funding from an NSERC Canadian Graduate Scholarship and C.R.F. is funded by an NSERC Postdoctoral Fellowship. P.R.P.-N., R.W.N., and J.C.C. were supported by NSERC Discovery grants. M.A. was funded by NSF RCN DEB-1840663. F.A. received funding from CAPES. M.T.K.A. was funded by CONICYT PIA APOYO CCTE AFB170008. J.R.B, T.C.L., and S.A.S were supported by Monmouth University School of Science Summer Research Program. E.G. was funded by Département de Biologie, Université de Moncton. C.G.-L. received funding from the Center of Applied Ecology and Sustainability (CAPES), and ANID PIA/BASAL FB0002. S.G. was funded by the Max Planck Society. P.J.-A. was funded by ANID PIA/BASAL FB210006. I.N. and M.S. were supported by Leiden Municipality. K.M.O. was funded by US NSF awards IOS-1557770 and DEB-1601641. J.C.P. thanks FAPESP process 2018/00107-3, and M.C.R. thanks CNPq and FAPESP. **Author contributions:** The project's lead team included B.C., C.R.F., S.G.I., M.T.J.J., S.K., L.S.M., S.M., R.W.N., P.R.P.-N., C.P., J.S.S., and A.T.T. M.T.J.J., R.W.N., and J.S.S. conceived of the project. H.T.F., M.T.J.J, J.S.S., and A.T.T. collected spatial environmental and city data. M.T.J.J., P.R.P.-N., and J.S.S. performed statistical analyses. R.W.N. and J.S.S. performed bioinformatic and genomic analyses. B.C., C.R.F., S.G.I., M.T.J.J., S.K., L.S.M., S.M., R.W.N., C.P., J.S.S., E.C., and J.M.-S. contributed reagents, materials, technical skills, or analysis tools. All remaining authors designed transects, collected samples, and analyzed data. M.T.J.J. and J.S.S. wrote the paper with input from the lead team; all authors provided comments on drafts of the paper. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** All code and environmental and phenotypic data are available on the GitHub page for J.S.S. (https://github.com/James-S-Santangelo/glue_pc) and additionally archived on Zenodo (*31*). BAM files have been deposited in the European Nucleotide Archive (ENA BioProject PRJEB48967).

[1]Department of Biology, University of Toronto Mississauga, Mississauga, ON, Canada. [2]Centre for Urban Environments, University of Toronto Mississauga, Mississauga, ON, Canada. [3]Department of Biology, University of North Carolina, Chapel Hill, NC, USA. [4]Department of Biology, University of Louisiana, Lafayette, LA, USA. [5]Department of Biology, Queen's University, Kingston, ON, Canada. [6]Department of Biology, Concordia University, Montreal, QC, Canada. [7]Department of Biological Sciences, DePaul University, Chicago, IL, USA. [8]Department of Biology, DePauw University, Greencastle, IN, USA. [9]Department of Urban Design and Planning, University of Washington, Seattle, WA, USA. [10]Colegio de Ciencias Biológicas y Ambientales, Universidad San Francisco de Quito USFQ, Quito, Ecuador. [11]Department of Genetics, University of Georgia, Athens, GA, USA. [12]Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden. [13]Field Science Center for Northern Biosphere, Hokkaido University, Sapporo, Hokkaido, Japan. [14]Natural History Museum, Zoology, University of New England, Armidale, NSW, Australia. [15]Programa de Pós-Graduação em Geografia da UFMT, campus de Rondonópolis, Cuiabá, Brazil. [16]Department of Botany and Biodiversity Research Centre, University of British Columbia, Vancouver, BC, Canada. [17]Graduate Program in Genome Sciences and Technology, Genome Sciences Centre, University of British Columbia, Vancouver, British Columbia, Canada. [18]Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia, Canada. [19]Red de Biología Evolutiva, Instituto de Ecología, A. C., Xalapa, Mexico. [20]School of the Environment, Yale University, New Haven, CT, USA. [21]Departamento de Ciencias Ecológicas, Facultad de Ciencias, Universidad de Chile, Santiago, Chile. [22]Instituto de Ecología y Biodiversidad, Universidad de Chile, Santiago, Chile. [23]Department of Biology, Mount Allison University, Sackville, NB, Canada. [24]Red de Ecoetología, Instituto de Ecología A. C., Xalapa, Mexico. [25]Department of Biology, University of Ottawa, Ottawa, ON, Canada. [26]Department of Zoology, University of Cambridge, Cambridge, UK. [27]Department of Biology, Washington University in St. Louis, St. Louis, MO, USA. [28]Department of Biology, University of Miami, Miami, FL, USA. [29]Centro de Investigación en Recursos Naturales y Sustentabilidad (CIRENYS), Universidad Bernardo O'Higgins, Santiago, Chile. [30]Department of Biology, University of La Verne, La Verne, CA, USA. [31]Département des sciences du bois et de la forêt, Université Laval, Quebec, QC, Canada. [32]Evolution & Ecology Research Centre, School of Biological, Earth and Environmental Sciences, UNSW Sydney, Sydney, NSW, Australia. [33]Department of Biology, Ghent University, Ghent, Belgium. [34]Department of Biology, Monmouth University, West Long Branch, NJ, USA. [35]Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, Lisboa, Portugal. [36]Department of Biology, KU Leuven, Leuven, Belgium. [37]School of Agriculture and Environment, Wildlife and Ecology group, Massey University, Palmerston North, Manawatu, New Zealand. [38]Department of Biological Sciences, University of Cape Town, Cape Town, South Africa. [39]Institute of Landscape Ecology, University of Münster, Münster, Germany. [40]Gosnell School of Life Sciences, Rochester Institute of Technology, Rochester, NY, USA. [41]Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada. [42]Louis Calder Center and Department of Biological Sciences, Fordham University, Armonk, NY, USA. [43]Departamento de Ecología Tropical, Universidad Autónoma de Yucatán, Mérida, Yucatán, México. [44]School of Life Sciences, University of Sussex, Brighton, UK. [45]Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm, Sweden. [46]iBIOTROP Instituto de Biodiversidad Tropical, Universidad San Francisco de Quito, Quito, Ecuador. [47]Department of Biology, San Francisco State University, San Francisco, CA, USA. [48]Unidad de Recursos Naturales, Centro de Investigación Científica de Yucatán AC, Mérida, Yucatán, México. [49]School of Ecological and Environmental Sciences, East China Normal University, Shanghai, China. [50]Shanghai Engineering Research Center of Sustainable Plant Innovation, Shanghai 200231, China. [51]Centre for Ecosystem Science, School of Biological, Earth and Environmental Sciences, UNSW Sydney, Sydney, NSW, Australia. [52]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA. [53]Department of Biosciences, Rice University, Houston, TX, USA. [54]IFEVA, Universidad de Buenos Aires, CONICET, Facultad de Agronomía, Buenos Aires, Argentina. [55]Biology Department, Saint Mary's University, Halifax, NS, Canada. [56]Department of Biological Sciences, Universidad de los Andes, Bogotá, Colombia. [57]Department of Biology and Biochemistry, University of Houston, Houston, TX, USA. [58]ECOBIO (Ecosystèmes, biodiversité, évolution), Université de Rennes, Rennes, France. [59]Department of Zoology and Biodiversity Research Centre, University of British Columbia, Vancouver, BC, Canada. [60]Department of Environmental Studies, Dordt University, Sioux Center, IA, USA. [61]Department of Biology, Minneapolis Community and Technical College, Minneapolis, MN, USA. [62]Department of Natural Sciences, Ecology and Environment Research Centre, Manchester Metropolitan University, Manchester, UK. [63]Instituto de Investigaciones en Ecosistemas y Sustentabilidad, UNAM, Morelia, Mexico. [64]Department of Botany, School of Biology, Aristotle University of Thessaloniki, Thessaloniki, Greece. [65]Faculty of Biological and Environmental Science, Organismal & Evolutionary Biology Research Programme, University of Helsinki, Helsinki, Finland. [66]Corporación Científica Ingeobosque, Medellín, Antioquia, Colombia. [67]GTA Colombia S.A.S. Envigado, Antioquia, Colombia. [68]Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, Scotland, UK. [69]Department of Biology, Hendrix College, Conway, AR, USA. [70]Department of Ecological Science, Vrije Universiteit Amsterdam, Amsterdam, Netherlands. [71]Departamento de Ciencias Biológicas y Agropecuarias, Universidad Técnica Particular de Loja, Loja, Ecuador. [72]Departamento de Biologia, Universidade Federal de Santa Maria (UFSM), Santa Maria, Rio Grande do Sul, Brazil. [73]Department of Plant Sciences, School of Biology, College of Science, University of Tehran, Tehran, Iran. [74]NTNU University Museum, Norwegian University of Science and Technology, 7491 Trondheim, Norway. [75]Red de Estudios Moleculares Avanzados, Instituto de Ecología A. C., Xalapa, Mexico. [76]School of Biological Sciences, University of Reading, Whiteknights Park, Reading, Berkshire, UK. [77]Department of Biology, Northern Arizona University, Flagstaff, AZ, USA. [78]Department of Biological Sciences, MacEwan University, Edmonton, AB, Canada. [79]Max Planck Institute for Plant Breeding Research, Cologne, Germany. [80]Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Ciudad de México, México. [81]Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany. [82]BIO5 Institute, University of Arizona, Tucson, AZ, USA. [83]Alaska Center for Conservation Science, University of Alaska Anchorage, Anchorage, AK, USA. [84]Tropical Diversity, Royal Botanical Garden of Edinburgh, Edinburgh, UK. [85]Départemente de biologie, Université de Moncton, Moncton, New Brunswick, Canada. [86]Department of Biological Sciences, University of Manitoba, Winnipeg, MB, Canada. [87]Departments of Microbiology & Statistics, University of Manitoba, Winnipeg, MB, Canada. [88]Department of Biology, University of New Brunswick, Fredericton, NB, Canada. [89]Department of Biology, Kalamazoo College, Kalamazoo, MI, USA. [90]BioProtection Research Centre, Lincoln University, Lincoln, Canterbury, New Zealand. [91]Departamento de Ciencias, Facultad de Artes Liberales, Universidad Adolfo Ibáñez, Santiago, Chile. [92]Department of Ecology, Evolution, and Behaviour University of Minnesota, Minneapolis, MN, USA. [93]Department of Biological Sciences, Brock University, St. Catharines, Ontario, Canada. [94]Department of Environmental Toxicology, University of California, Davis, CA, USA. [95]ICB - University of Talca, Chile. [96]School of Molecular and Life Science, Curtin University, Perth, Australia. [97]College of Science, Health, Engineering and Education, Murdoch University, Murdoch, WA, Australia. [98]School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia. [99]School of Biological Sciences, Monash University, Melbourne, VIC, Australia. [100]Department of Biological Sciences, Wayne State University, Detroit, MI, USA. [101]Department of Biology, Western Oregon University, Monmouth, OR, USA. [102]School of Natural Resources and the Environment, University of Arizona, Tucson, AZ, USA. [103]Departamento de Ecología Humana, Cinvestav Mérida, Yucatan, México. [104]Departamento de Ciencias Biológicas y Departamento de Ecología y Biodiversidad, Facultad de Ciencias de la Vida, Universidad Andrés Bello, Santiago, Chile. [105]Institute of Ecology and Biodiversity (IEB), Chile. [106]Department of Biology, Lund University, Lund, Sweden. [107]Department of Biology, Norwegian University of Science and Technology, Trondheim, Norway. [108]Escuela Superiro de Desarrollo Sustentable, Universidad Autónoma de Guerrero -CONACYT, Las Tunas, Mexico. [109]Clarkson Secondary School, Peel District School Board, Mississauga, ON, Canada. [110]Homelands Sr. Public School, Peel District School Board, Mississauga, ON, Canada. [111]Department of Biological Sciences, University of Illinois at Chicago, Chicago, IL, USA. [112]St. James Catholic Global Learning Centre, Dufferin-Peel Catholic District School Board, Mississauga ON, Canada. [113]Department of Biosciences, University of Calgary, Calgary, AB, Canada. [114]Ecological Processes Branch, U.S. Army ERDC-CERL, Champaign, IL, USA. [115]Department of Biology, Oberlin College, Oberlin, OH, USA. [116]Escuela Nacional de Estudios Superiores Unidad Morelia, UNAM, Morelia, Mexico. [117]Institute of Evolution and Ecology, University of Tübingen, Tübingen, Germany. [118]Department of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse, Zurich, Switzerland. [119]Urban Wildlife Institute, Department of Conservation and Science, Lincoln Park Zoo, Chicago, IL, USA. [120]Departamento de Ecología, Universidad Católica de la Santísima Concepción, Concepción, Chile. [121]Department of Biological Sciences, University of Denver, Denver, CO, USA. [122]Department of Biological Sciences, Mississippi State University, Starkville, MS, USA. [123]Department of Biology, Center for Computational & Integrative Biology, Rutgers University-Camden, Camden, NJ, USA. [124]Programa de Pós-Graduação em Geografia da UFMT, campus de Rondonópolis, Brasil. [125]Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, China. [126]Department of Chemistry & Biochemistry, Laurentian University, Sudbury, ON, Canada. [127]Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing, China. [128]School of BioSciences, University of Melbourne, Melbourne, VIC, Australia. [129]Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México, Coyoacán, Mexico City, 04510, Mexico. [130]Department of Biological Sciences, Auburn University, Auburn, AL, USA. [131]Department of Entomology and Nematology, University of California, Davis, CA, USA. [132]Department of Biology, University of New Mexico, Albuquerque, NM, USA. [133]Department of Biology, University of Wisconsin - Eau Claire, Eau Claire, WI 54701. [134]Agriculture Institute, Iranian Research Organization for Science and Technology (IROST), Tehran, Iran. [135]Department of Biology, Colby College, Waterville, ME, USA. [136]Instituto de Biología, Universidad de Antioquia, Medellín, Colombia. [137]Department of Biology, University of Massachusetts Boston, Boston, MA, USA. [138]Agricultural Biology, Colorado State University, Fort Collins, CO, USA. [139]Departamento de Biología Vegetal y Ecología, Facultad de Biología, Universidad de Sevilla, Av. Reina Mercedes s/n, 41012

Sevilla, Spain. [140]Facultad de Estudios Interdisciplinarios, Centro GEMA- Genómica, Ecología y Medio Ambiente, Universidad Mayor, Santiago, Chile. [141]Evolutionary Ecology Group, Naturalis Biodiversity Center, Leiden, Netherlands. [142]Department of Biology and Chemistry, Nipissing University, North Bay, ON, Canada. [143]Center for Ecological Research, Kyoto University, Otsu, Shiga, Japan. [144]Bonanza Creek Long Term Ecological Research Program, University of Alaska Fairbanks, Fairbanks, AK, USA. [145]Department of Botany and Molecular Evolution, Senckenberg Research Institute and Natural History Museum Frankfurt, Frankfurt am Main, Germany. [146]Departamento de Biodiversidade, Instituto de Biociências, Univ Estadual Paulista - UNESP, Rio Claro, São Paulo, Brazil. [147]Nelson Institute for Environmental Studies, University of Wisconsin-Madison, Madison, WI, USA. [148]Department of Biology, California State University, Northridge, Los Angeles, CA, USA. [149]Department of Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden. [150]Facultad de Ciencias y Biotecnologia, Universidad CES, Medellín, Colombia. [151]Department of Biology, Hofstra University, Long Island, NY, USA. [152]Faculty of Biosciences and Aquaculture, Nord University, Bodø, Norway. [153]Division of Biological Sciences, University of California San Diego, San Diego, CA, USA. [154]Department of Biology, University of Richmond, Richmond, VA, USA. [155]Estación de Biodiversidad Tiputini, Colegio de Ciencias Biológicas y Ambientales, Universidad San Francisco de Quito USFQ, Quito, Ecuador. [156]Department of Biological Sciences, Institute of Environment, Florida International University, Miami, FL, USA. [157]Agronomy Department, University of Almería, Almería, Spain. [158]Department of Biological Sciences and Center for Urban Ecology and Sustainability, Butler University, Indianapolis, IN, USA. [159]Department of Biological Sciences, Louisiana State University, Baton Rouge, LA, USA. [160]Faculty of Biological Sciences, Goethe University Frankfurt, Frankfurt am Main, Germany. [161]Institute of Biology Leiden, Leiden University, Leiden, Netherlands. [162]Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland. [163]Department of Biology, University of Louisville, Louisville, KY, USA. [164]Organization for Programs on Environmental Science, University of Tokyo, Tokyo, Japan. [165]Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique et Evolution, 91405, Orsay, France. [166]Department of Biology, Providence College, Providence, RI, USA. [167]General Zoology, Institute for Biology, Martin Luther University Halle-Wittenberg, Halle, Germany. [168]International Arctic Research Center, University of Alaska Fairbanks, Fairbanks, AK, USA. [169]Science, Technology and Society Department, Rochester Institute of Technology, Rochester, NY, USA. [170]SLU Swedish Species Information Centre, Swedish University of Agricultural Sciences, Uppsala, Sweden. [171]Department of Biology, Westfield State University, Westfield, MA, USA. [172]Centre of New Technologies, University of Warsaw, Warsaw, Poland. [173]Department of Biology, Stanford University, Stanford, CA, USA. [174]UMR 0980 BAGAP, Agrocampus Ouest-ESA-INRA, Rennes, France. [175]Plant Biology Department, Michigan State University, East Lansing, MI, USA. [176]Biology Department, Davidson College, Davidson, NC, USA. [177]College of Horticulture and Forestry Sciences/ Hubei Engineering Technology Research Center for Forestry Information, Huazhong Agricultural University, Wuhan, China, Hubei, China. [178]School of Life Sciences, Technical University of Munich, Munich, Germany. [179]School of Life Sciences, Lanzhou University, Lanzhou, China. [180]Institute of Ecology and Evolution, University of Bern, Bern, Switzerland. [181]Department of Evolution, Ecology and Behaviour, University of Liverpool, Liverpool, UK.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abk0989
Material and Methods
Supplementary Text
Figs. S1 to S16
Tables S1 to S15
References (*32*–*125*)
MDAR Reproducibility Checklist

21 July 2021; accepted 11 February 2022
10.1126/science.abk0989

MOLECULAR BIOLOGY

The histone H3.1 variant regulates TONSOKU-mediated DNA repair during replication

Hossein Davarinejad[1]†, Yi-Chun Huang[2]†, Benoit Mermaz[2], Chantal LeBlanc[2], Axel Poulet[2], Geoffrey Thomson[2], Valentin Joly[2], Marcelo Muñoz[3,4], Alexis Arvanitis-Vigneault[1], Devisree Valsakumar[5,6], Gonzalo Villarino[2], Alex Ross[3,4], Benjamin H. Rotstein[4,7], Emilio I. Alarcon[3,4], Joseph S. Brunzelle[8], Philipp Voigt[5,6], Jie Dong[2,9], Jean-François Couture[1]*, Yannick Jacob[2]*

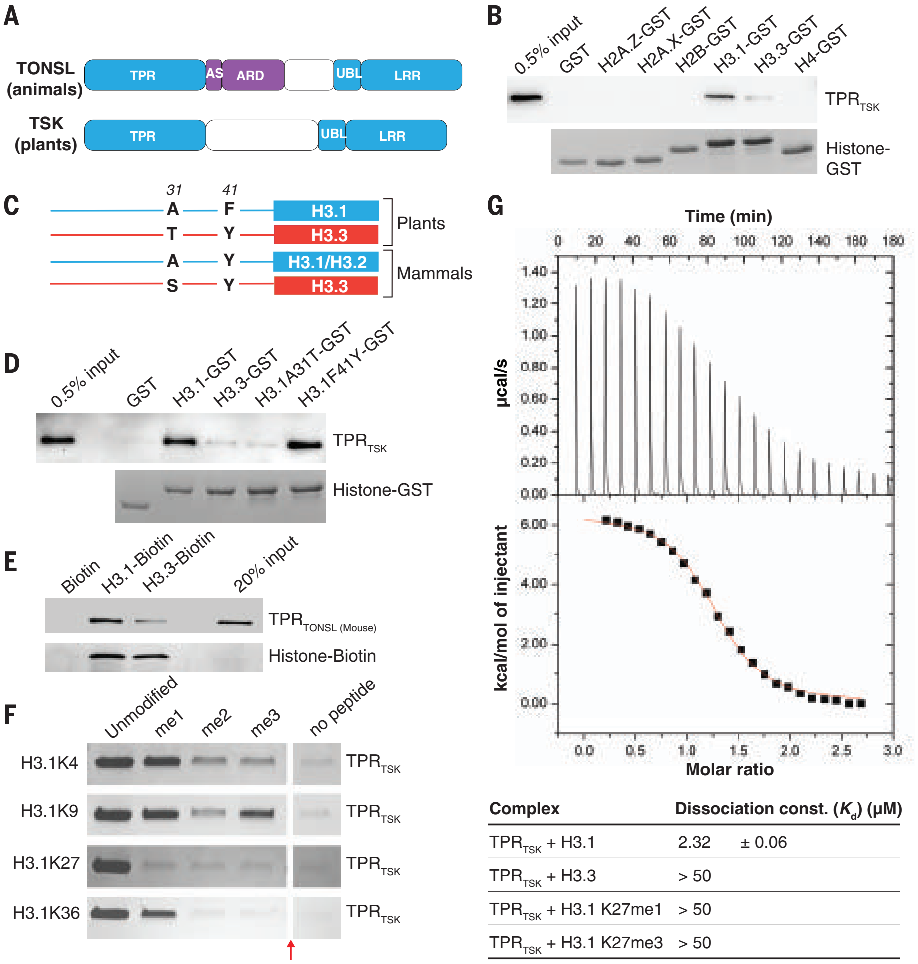
The tail of replication-dependent histone H3.1 varies from that of replication-independent H3.3 at the amino acid located at position 31 in plants and animals, but no function has been assigned to this residue to demonstrate a unique and conserved role for H3.1 during replication. We found that TONSOKU (TSK/TONSL), which rescues broken replication forks, specifically interacts with H3.1 via recognition of alanine 31 by its tetratricopeptide repeat domain. Our results indicate that genomic instability in the absence of ATXR5/ATXR6-catalyzed histone H3 lysine 27 monomethylation in plants depends on H3.1, TSK, and DNA polymerase theta (Pol θ). This work reveals an H3.1-specific function during replication and a common strategy used in multicellular eukaryotes for regulating post-replicative chromatin maturation and TSK, which relies on histone monomethyltransferases and reading of the H3.1 variant.

Chromatin replication requires multiple regulatory mechanisms to ensure the maintenance of genome integrity. One of these mechanisms relies on TONSOKU-LIKE (TONSL), a key player in initiating homologous recombination (HR) when replication forks encounter double-stranded DNA breaks (DSBs) (*1*–*8*). In animals, TONSL is recruited to chromatin via its ankyrin repeat domain (ARD), which specifically interacts with unmethylated histone H4 Lys^{20} (H4K20me0) (*1*, *9*). Post-replicative maturation of chromatin is accomplished via SET8/PR-Set7/SETD8 (*10*–*12*), which monomethylates histone H4 Lys^{20} (H4K20me1) and thus prevents TONSL from binding chromatin and initiating HR-based DNA repair outside of DNA replication and the G_2 phase of the cell cycle (*9*). Comparative analysis shows that plants contain a TONSL ortholog (TSK/BRUSHY1/MGOUN3) (*7*, *8*, *13*) but lack SET8. In addition, the ARD of TONSL in animals is not conserved in TSK orthologs (Fig. 1A) (*1*), indicating that post-replicative chromatin maturation in plants is unlikely to depend on the methylated state of H4 Lys^{20}.

We reasoned that TSK might interact with histones in plants through a different domain. Sequence alignment of TSK orthologs shows extensive similarity in the N-terminal tetratricopeptide repeat (TPR) domain (fig. S1), which is conserved in animals (Fig. 1A) (*1*, *4*). Many TPR domains bind long peptides (>20 amino acids), adopting an extended conformation (*14*). We therefore hypothesized that one of the N-terminal unstructured tails of histones could specifically interact with the TPR domain of TSK (TPR_{TSK}). To assess this, we performed in vitro binding assays with *Arabidopsis thaliana* TPR_{TSK} and the tails of different histones. We detected binding of TPR_{TSK} with H3 variants, with stronger binding for H3.1 relative to H3.3 (Fig. 1B). A preference for TPR_{TSK} to bind H3.1 over H3.3 was also observed using nucleosomes and in *A. thaliana* protoplasts (fig. S2, A to C). In vascular plants, amino acids 31 and 41 vary between the N-terminal tails of H3.1 and H3.3 (Fig. 1C) (*15*). We created hybrid H3.1/$\text{H3.3}_{\text{tail}}$-GST fusion proteins based on these differences and determined that only alanine at position 31 of H3.1 (H3.1 Ala^{31}) is required for the H3.1-binding specificity of TPR_{TSK} (Fig. 1D). Variation at position 31 of H3 is also observed between replication-dependent H3.1/H3.2 variants and the replication-independent H3.3 in mammals (Fig. 1C). Similarly to plant TSK orthologs, the TPR domain of mouse TONSL also interacts preferentially with H3.1 relative to H3.3 (Fig. 1E). We then assessed the impact of TPR_{TSK} binding to H3.1 in the context of methylation at different lysine residues in the N-terminal tail of H3.1. We found that increasing levels of methylation at Lys^{4}, Lys^{9}, Lys^{27}, and Lys^{36} have a negative impact on the interaction of TPR_{TSK} and H3.1, with binding being

---

[1]Ottawa Institute of Systems Biology, Department of Biochemistry, Microbiology and Immunology, Faculty of Medicine, University of Ottawa, Ottawa, Ontario K1H 8M5, Canada. [2]Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06511, USA. [3]Wellcome Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, UK. [4]Epigenetics Programme, Babraham Institute, Cambridge CB22 3AT, UK. [5]BEaTS Research Laboratory, Division of Cardiac Surgery, University of Ottawa Heart Institute, Ottawa, Ontario K1Y 4W7, Canada. [6]Department of Biochemistry, Microbiology, and Immunology, Faculty of Medicine, University of Ottawa, Ottawa, Ontario K1H 8M5, Canada. [7]University of Ottawa Heart Institute, Ottawa, Ontario K1Y 4W7, Canada. [8]Department of Molecular Pharmacology and Biological Chemistry, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA. [9]Institute of Crop Science, Zhejiang University, Hangzhou 310058, China.
***Corresponding author. Email: yannick.jacob@yale.edu (Y.J.); jean-francois.couture@uottawa.ca (J.-F.C.)**
†These authors contributed equally to this work.



Complex	Dissociation const. (K_d) (μM)
TPR$_{TSK}$ + H3.1	2.32 ± 0.06
TPR$_{TSK}$ + H3.3	> 50
TPR$_{TSK}$ + H3.1 K27me1	> 50
TPR$_{TSK}$ + H3.1 K27me3	> 50

Fig. 1. The TPR domain of TSK specifically interacts with the N-terminal tail of the H3.1 variant. (**A**) Domain architecture of animal and plant TONSL/TSK. TPR, tetra tricopeptide repeats; AS, acidic sequence; ARD, ankyrin repeat domain; UBL, ubiquitin-like; LRR, leucine-rich repeats. Conserved domains are shown in blue. (**B**) Pull-down assay using TPR$_{TSK}$ and GST (glutathione *S*-transferase) tagged with the N-terminal tails of histones H2A.Z, H2A.X, H2B, H3.1, H3.3, and H4 from plants. (**C**) Representation of plant and mammalian H3.1/H3.2 (blue) and H3.3 (red) H3 variants. Thin lines and blocks represent the histone tails and cores, respectively; numbers indicate amino acid positions in H3. (**D**) Peptide pull-down assay using plant TPR$_{TSK}$ and GST tagged with the tails of histones H3.1, H3.3, H3.1A31T, and H3.1F41Y. (**E**) Peptide pull-down assay using mouse TPR$_{TONSL}$ and biotin-tagged histones H3.1 and H3.3 (full-length proteins) from mammals. (**F**) Peptide pull-down assay using plant TPR$_{TSK}$ and methylated peptides at Lys^{4}, Lys^{9}, Lys^{27}, and Lys^{36} of H3.1 (amino acids 1 to 45). The red arrow indicates a gel lane that was removed. (**G**) Isothermal titration calorimetry assay using plant TPR$_{TSK}$ and different H3 peptides.

most sensitive to methylation at Lys^{27} (Fig. 1, F and G, and fig. S2D). The binding profile of TSK on histone H3 suggests a preference for binding the newly synthesized H3.1 variant.

To gain mechanistic insights into how TSK discriminates H3.1 from H3.3, we solved the crystal structure of the TPR$_{TSK}$-H3.1$_{(1-45)}$ complex at 3.17 Å resolution by using the TSK ortholog from *Citrus unshiu* (CuTSK) (fig. S1 and table S1). TPR$_{TSK}$ folds as 11 TPR motifs placed in tandem, which collectively form a hollow solenoid tube (Fig. 2, A and B, and fig. S3A). The C-terminal lobe of the tube is composed of TPRs 6 to 11 and generates a wide channel in which two segments of H3.1 (Lys^{4} to Lys^{9} and Lys^{18} to Ala^{24}) are found along opposite sides of its wall (Fig. 2, A and C). In the center lobe, TPRs 3 to 7 form a narrow tunnel that encircles Ala^{25} to Pro^{30} of H3.1 (Fig. 2, A and C). H3.1 Lys^{27} is located inside a polar pocket, where its ε-amine is surrounded by the side chains of Asp^{234}, Cys^{238}, and Ser^{208} and the backbone carbonyl groups of Asp^{234} and Gly^{246} (Fig. 2D). The polarity of this pocket makes it nonconducive for the binding of hydrophobic moieties such as methyl groups, thus explaining the large decrease in binding affinity of TPR$_{TSK}$ to H3.1 when Lys^{27} is mono- or trimethylated (Fig. 1G and fig. S2D). TPRs 1 to 3 make up the N-terminal lobe of TPR$_{TSK}$, which forms an open channel that accommodates Pro^{30} to Arg^{40} of H3.1 (Fig. 2, A and C). A deep pocket formed between α helices 2 to 4 (TPRs 1 and 2) is occupied by the side chain of H3.1 Lys^{36}, where its ε-amine is in close proximity to the carboxyl group of Asp^{54} (Fig. 2, A, C, and E). The side chain of H3.1 Ala^{31} is oriented toward the aliphatic portion of three residues (Arg^{109}, Gln^{113}, and Gln^{72}) strictly conserved among plant TSK orthologs (Fig. 2F and fig. S3A). These residues form a shallow pocket in which Gln^{113} and Gln^{72} also likely interact with the H3.1 backbone via hydrogen bonds with the carbonyl group of Gly^{34} and the amide group of Ala^{31}, respectively (Fig. 2, F and G). Consistent with our binding assays (Fig. 1, B, D, and G, and fig. S2D), modeling of an A31T substitution (corresponding to H3.3) in H3.1 generates van der Waals clashes between the Cγ methyl group of Thr^{31} and the aliphatic chain of Gln^{113}, and similarly between the hydroxyl group of Thr^{31} and Arg^{109} (fig. S3B). We mutated various amino acids of TPR$_{TSK}$ from different H3.1 binding pockets and validated that they contribute to the TPR$_{TSK}$-H3.1 interaction (fig. S3C). Overall, the structure of the TPR$_{TSK}$-H3.1 complex supports our finding that TSK preferentially binds the replication-dependent H3.1 variant.

In plants, the histone H3 Lys^{27} monomethyltransferases ATXR5 and ATXR6 (ATXR5/6) maintain genome stability by specifically methylating the H3.1 variant (H3.1K27me1) during DNA replication (*16*–*18*). Loss of H3.1K27me1 in *atxr5/6* double mutants results in genomic amplification of heterochromatin, transposable element derepression, and disruption of heterochromatin structure (*17*, *19*). Additional work has shown that heterochromatin amplification in *atxr5/6* mutants is dependent on DNA repair (*20*). Therefore, ATXR5/6 may play a role analogous to the mammalian H4K20 monomethyltransferase SET8 in regulating TONSL/TSK activity, with the difference that H3.1K27me1, not H4K20me1, is the key histone modification used in plants to prevent TSK from interacting with chromatin and initiating DNA repair. To validate this model, we generated an *atxr5/6 tsk* triple mutant in *A. thaliana* (fig. S4A). Flow cytometry analyses of *atxr5/6 tsk* mutants showed suppression of heterochromatin amplification induced by the absence of H3.1K27me1, as represented by the loss of the broad peaks corresponding to 8C and 16C endoreduplicated nuclei in *atxr5/6* mutants (Fig. 3A and fig. S4B).

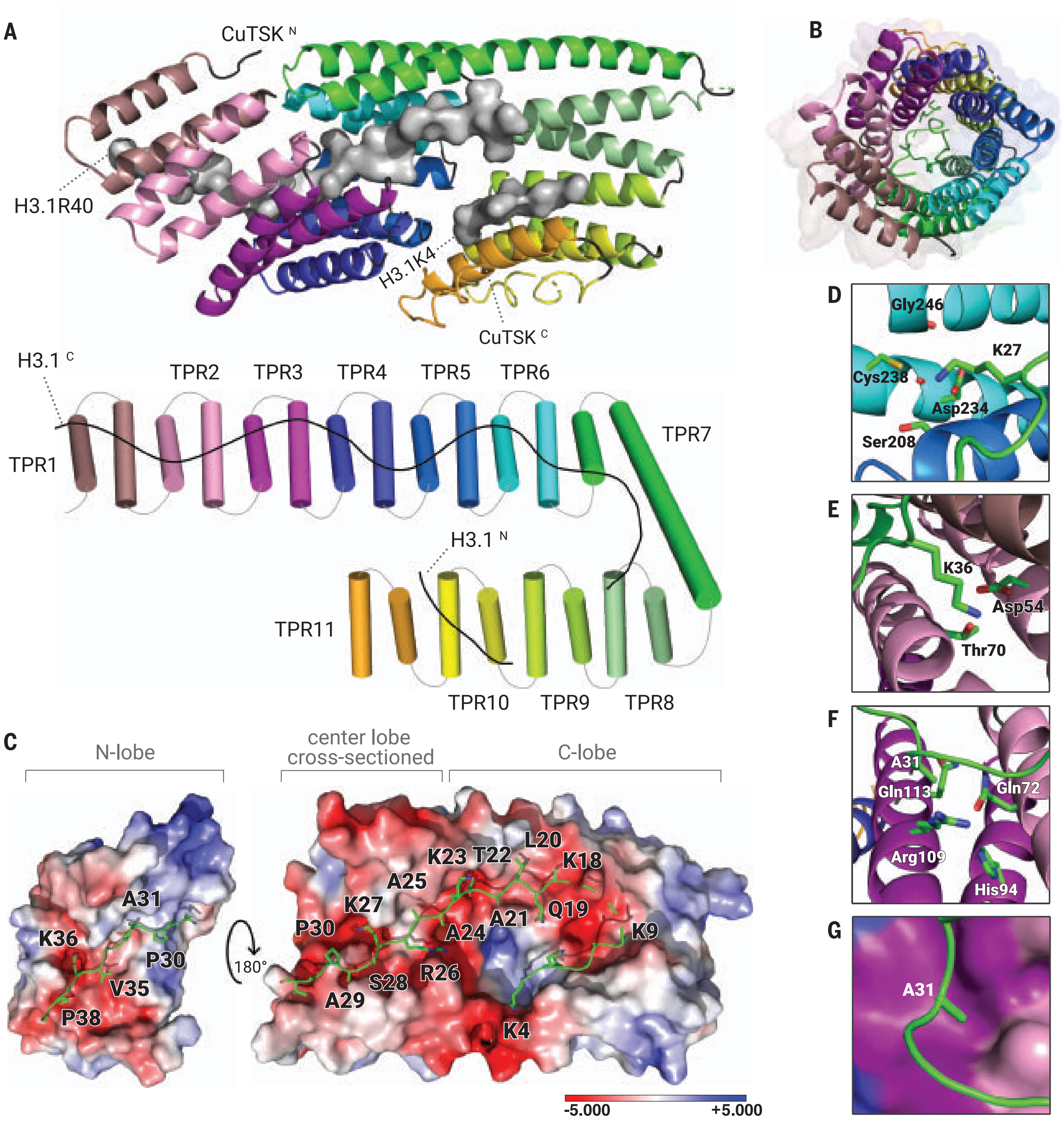


Fig. 2. Crystal structure of plant TPR$_{TSK}$ bound to the H3.1 tail. (**A**) The TPR domain is depicted as a cartoon (top) or a cylinder (bottom) with individual TPR motifs as distinct colors. H3.1 is shown as surfaces (top) or line (bottom). (**B**) Channel view of the TPR solenoid tube showing the space inside the tube where H3.1 is extended (represented as a green line). (**C**) Surface representation of the TPR domain shown as electrostatic potential gradients contoured from +5.000 $k_BT\ e^{-1}$ (blue) to −5.000 $k_BT\ e^{-1}$ (red), where e is the electron, T is absolute temperature, and k_B is the Boltzmann constant. H3.1 is depicted as sticks. The N-terminal lobe (N-lobe) is rotated 180° along the horizontal axis relative to the center lobe and the C-terminal lobe (C-lobe). The surface of the center lobe is sectioned off to reveal the underlying segment of H3.1. (**D** to **F**) Amino acid residues from TPR$_{TSK}$ (3-letter code) interacting with H3.1 residues (1-letter code) in their binding pockets are shown for Lys^{27} (D), Lys^{36} (E), and Ala^{31} (F). (**G**) Surface representation of the H3.1A31 binding pocket. Surface colors correspond to those of TPR helices shown in (F).

This result was confirmed by genome sequencing of 16C nuclei from leaf tissue (Fig. 3B). We also observed transcriptional suppression of the genome instability marker *BRCA1*, which is highly expressed in *atxr5/6* but not in *atxr5/6 tsk* (fig. S4C) (*21*). In addition, the number of chromocenters adopting a hollowed sphere conformation characteristic of *atxr5/6* mutants was decreased when *TSK* was inactivated (Fig. 3, C and D) (*20*). Similarly, transcriptional derepression in heterochromatin of *atxr5/6* mutants was reduced in the absence of *TSK* (Fig. 3E, fig. S4, D to F, and table S2). These results indicate that the heterochromatic defects caused by the loss of H3.1K27me1 in plants are dependent on TSK.

In mammals, TONSL is recruited to newly replicated chromatin and promotes DNA repair via HR at broken replication forks (*1*, *3*–*6*, *9*). Cell cycle expression analysis in synchronized tobacco cells indicates that *TSK*

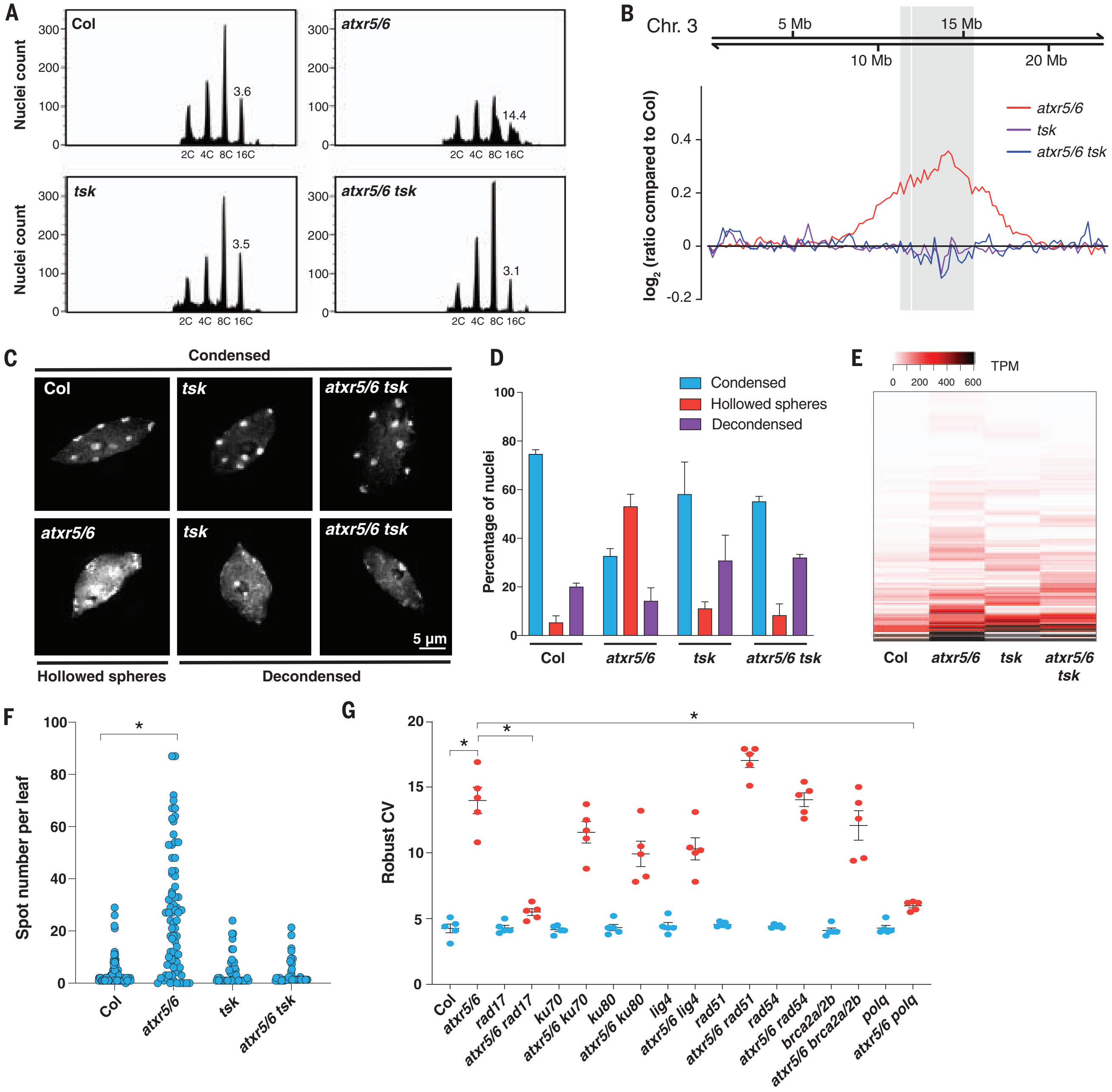


Fig. 3. Mutations in *TSK* suppress heterochromatin amplification of *atxr5/6* mutants. (**A**) Flow cytometry profiles of Col, *atxr5/6*, *tsk*, and *atxr5/6 tsk* leaf nuclei. The numbers below the peaks indicate ploidy levels of the nuclei. The numbers above the 16C peaks indicate the robust coefficient of variation (rCV). (**B**) Chromosomal view (chromosome 3 of *A. thaliana*) of DNA sequencing reads from sorted 16C nuclei. The pericentromeric region is highlighted in gray. (**C**) Leaf interphase nuclei of Col, *atxr5/6*, *tsk*, and *atxr5/6 tsk* stained with DAPI. (**D**) Quantification of nuclei from the experiment shown in (C). Error bars indicate SEM. (**E**) Heatmap showing the relative expression levels of *atxr5/6*-induced transposable elements as measured by TPM (transcripts per million). (**F**) Average number of blue spots per leaf in Col, *atxr5/6*, *tsk*, and *atxr5/6 tsk* as determined using a *GUS* reporter for homologous recombination. $^*P < 0.0001$ [Welch's analysis of variance (ANOVA) followed by Dunnett's T3 test]. (**G**) rCV values for 16C nuclei obtained by flow cytometry analyses. Each dot represents an independent biological replicate. Horizontal bars indicate means. Error bars represent SEM. $^*P < 0.05$ (Welch's ANOVA followed by Dunnett's T3 test).

is specifically expressed in S phase (*22*), which supports a conserved role for TSK during replication. To assess whether H3.1K27me1 suppresses HR activity in plants, we used a reporter system for HR based on intrachromosomal recombination restoring activity at a colorimetric *GUS* transgene (*23*). Our results show that *GUS* activity is much stronger in *atxr5/6* mutants than in wild-type plants, but not in *atxr5/6 tsk* mutants (Fig. 3F and fig. S5), thus indicating a role for H3.1K27me1 in preventing TSK-mediated HR in plants.

The protein kinases ATM and ATR, which participate in the early signaling steps leading to HR-mediated DNA repair (*24*), were previously shown to be required for inducing heterochromatin amplification in *atxr5/6* mutants (*20*). We therefore tested the contributions of different DNA repair pathways to the phenotypes observed in *atxr5/6* mutants. Mutating nonhomologous end-joining (*Ku70*, *Ku80*, and *LIG4*) or HR (*RAD51*, *RAD54*, and *BRCA2A/BRCA2B*) genes did not have a major effect on heterochromatin amplification in *atxr5/6*

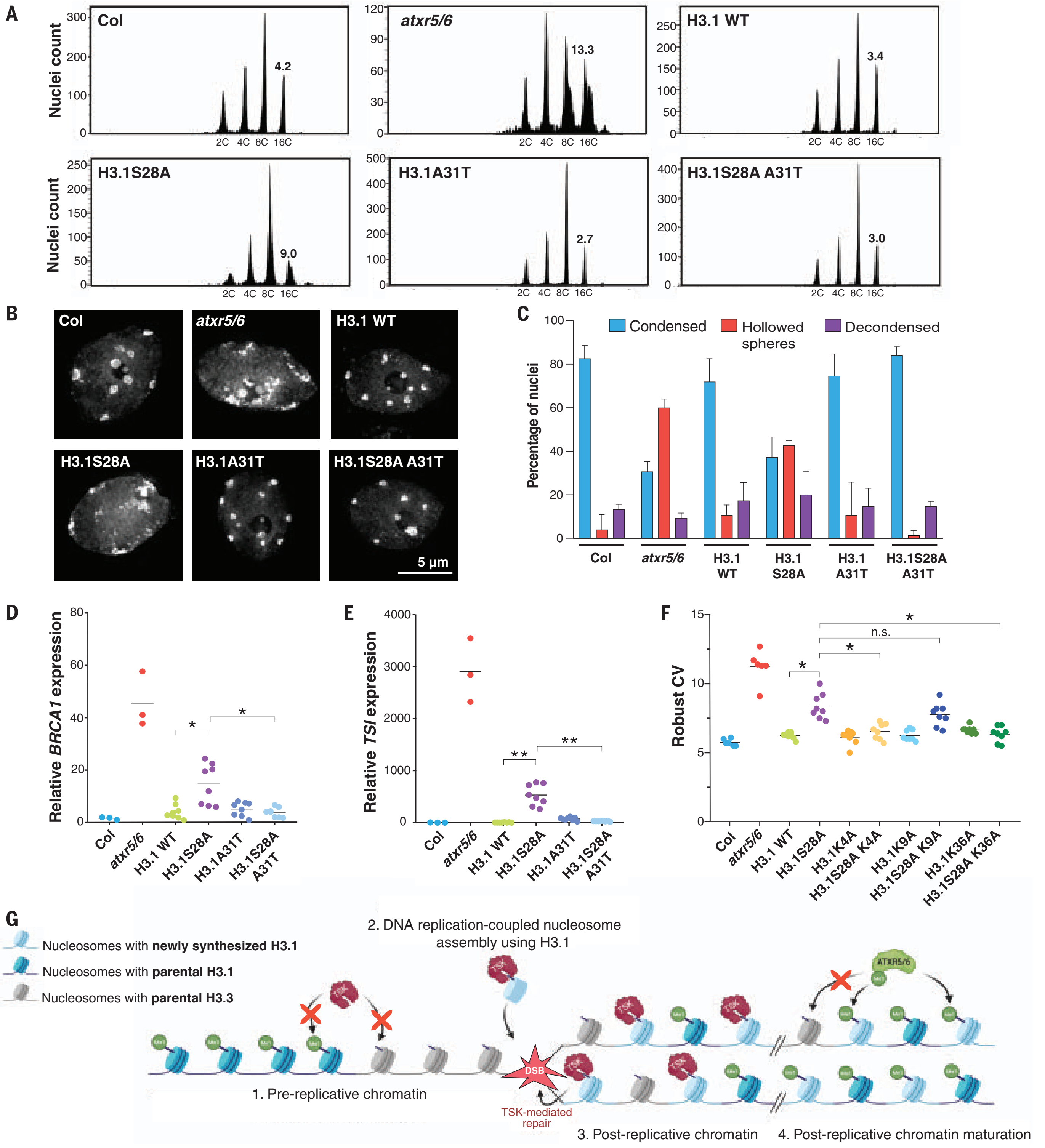


Fig. 4. H3.1 is required to mediate genomic instability in *atxr5/6* mutants. (**A**) Flow cytometry of leaf nuclei. Numbers below the peaks indicate ploidy, and those above indicate rCV. (**B**) Leaf nuclei of Col, *atxr5/6*, and first-generation (T_1) H3.1 lines stained with DAPI. (**C**) Quantification from nuclei in (B). Error bars indicate SEM. (**D** and **E**) Reverse transcription quantitative polymerase chain reaction of *BRCA1* and *TSI*. Horizontal bars indicate means. *$P < 0.05$, **$P < 0.001$ (Welch's ANOVA followed by Dunnett's T3 test). (**F**) rCV for 16C nuclei obtained by flow cytometry. For Col and *atxr5/6*, each dot represents a biological replicate. For the H3.1 lines, each dot represents one T_1 plant. Horizontal bars indicate means. *$P < 0.05$ (Welch's ANOVA followed by Dunnett's T3 test); n.s., not significantly different. (**G**) Model depicting the interplay among H3.1, TSK, and ATXR5/6 during replication. Step 1: TSK cannot interact with chromatin containing H3.3K27me0 or H3.1K27me1. Step 2: Newly synthesized copies of H3.1 (H3.1K27me0) in complex with TSK are inserted at replication forks. Step 3: DSBs caused by broken replication forks are repaired by TSK. Step 4: Monomethylation of newly inserted H3.1 (but not H3.3) at Lys^{27} by ATXR5/6 prevents binding of TSK.

mutants (Fig. 3G), although eliminating the HR recombinase RAD51 enhanced the morphological phenotypes of *atxr5/6* mutants (fig. S6A). In contrast, a mutation in *RAD17* suppressed heterochromatin amplification, loss of chromatin structure, and transcriptional derepression in *atxr5/6* mutants (Fig. 3G and fig. S6, B and C). RAD17 is responsible for loading the MRE11-RAD50-NBS1 complex that mediates DNA resection, one of the initial steps of HR (*25*). In animals, DNA resection can also lead to substrates that are repaired in an error-prone manner by Pol θ via polymerase theta–mediated end joining (TMEJ), which can create large tandem duplications of 1 kb to 1 Mb (*26*). We introduced a mutant allele of the *A. thaliana POLQ/TEBICHI* gene coding for Pol θ in the *atxr5/6* background and observed strong suppression of heterochromatin amplification and related phenotypes (Fig. 3G and fig. S7, A to D). Taken together, these results show that genomic instability in *atxr5/6* mutants is caused by a TSK-dependent pathway involving TMEJ.

The specificity of ATXR5/6 and TSK for replication-dependent H3.1 led us to hypothesize that this H3 variant is responsible for inducing TSK-mediated genomic instability in *atxr5/6* mutants. To test whether Ala^{31} of H3.1 is required for heterochromatin amplification in the absence of H3.1K27me1, we used a genetic system based on expression of the H3.1 point mutant H3.1S28A that mimics the phenotypes of *atxr5/6* mutants (Fig. 4, A to F, and fig. S8A) (*27*). The S28A point mutation prevents H3.1 Lys^{27} monomethylation by ATXR5/6 (*27*) but does not affect the binding of TPR$_{TSK}$ to H3.1 (fig. S8B), thus supporting a role for H3.1K27me1 in preventing the interaction of TSK with H3.1 in vivo. We then transformed *A. thaliana* with a transgene expressing H3.1S28A A31T (Ala^{31} replaced with Thr, as in plant H3.3 variants) and observed suppression of the heterochromatin phenotypes (Fig. 4, A to E, and fig. S8A), which demonstrates the importance of H3.1 Ala^{31} in regulating TSK activity in plants. The dependence of TSK on H3.1 explains why plants expressing H3.1A31T do not induce heterochromatin amplification despite losing ATXR5/6-catalyzed H3.1K27me1 (Fig. 4A) (*16*). A role for H3.1 Ala^{31} in mediating TSK activity is also supported by the observation that plants expressing H3.1A31T are hypersensitive to genotoxic stress, similarly to *tsk* and *h3.1* mutants (fig. S9, A to F).

We also used the H3.1S28A genetic system to assess the role of Lys^{4}, Lys^{9}, and Lys^{36} of H3.1 in contributing to the interaction with the TPR domain of TSK. Our in vivo results show that alanine replacement at Lys^{4} and Lys^{36} almost completely suppresses genomic instability and transcriptional derepression mediated by expression of H3.1S28A (Fig. 4F and fig. S10, A to D). These results are in line with in vitro experiments showing that H3.1K4A and H3.1K36A, but not H3.1K9A, strongly disrupt binding of TSK to H3.1 (fig. S10E). Finally, we used CRISPR to create a septuple mutant background, where all five *H3.1* genes are inactivated, in addition to mutations in *ATXR5/6* (*atxr5/6 h3.1*) (fig. S11, A and B). In *atxr5/6 h3.1* septuple mutants, both heterochromatin amplification and transcriptional derepression are suppressed (fig. S11, C and D), thus confirming that the H3.1 variant is required to induce these phenotypes. These results support the idea that TSK makes specific interactions with the N-terminal tail of the H3.1 variant in vivo to disrupt heterochromatin stability and silencing when H3.1K27me1 deposition is impaired.

Overall, this work uncovers a role for the TPR domain of TSK in selectively interacting with the H3.1 variant. Previous work in human cell lines has shown that the TSK ortholog TONSL copurifies with H3.1 in affinity purification/biochemical fractionation assays (*28*), and that TONSL-mediated double-stranded DNA break repair depends on the H3.1 chaperone CAF-1 (*2*). These findings, combined with our identification of the TPR domain of TSK/TONSL acting as an H3.1 reader, point to a model where post-replicative chromatin maturation in plants and animals relies on similar mechanisms involving H3.1 and clade-specific enzymes that monomethylate histones to prevent TSK/TONSL binding (Fig. 4G). In plants, monomethylation occurs at H3.1 Lys^{27} via ATXR5/6 and prevents binding of TSK through the TPR domain. In animals, SET8-mediated monomethylation at H4 Lys^{20} interferes with TONSL binding via the ARD domain (*9*). However, in both plants and animals, recruitment of TSK/TONSL to chromatin likely relies on the ability of the conserved TPR domain to preferentially interact with the H3.1 variant. Thus, our work reveals the importance of selectively incorporating H3.1 variants during DNA replication, as it confers a window of opportunity during the cell cycle for the TSK/TONSL DNA repair pathway to resolve broken replication forks.

REFERENCES AND NOTES

1. E. Duro *et al*., *Mol. Cell* **40**, 632–644 (2010).
2. T. H. Huang *et al*., *Mol. Cell* **69**, 879–892.e5 (2018).
3. B. C. O'Connell *et al*., *Mol. Cell* **40**, 645–657 (2010).
4. L. O'Donnell *et al*., *Mol. Cell* **40**, 619–631 (2010).
5. W. Piwko *et al*., *EMBO J.* **35**, 2584–2601 (2016).
6. W. Piwko *et al*., *EMBO J.* **29**, 4210–4222 (2010).
7. T. Suzuki *et al*., *Plant J.* **38**, 673–684 (2004).
8. S. Takeda *et al*., *Genes Dev.* **18**, 782–793 (2004).
9. G. Saredi *et al*., *Nature* **534**, 714–718 (2016).
10. J. Fang *et al*., *Curr. Biol.* **12**, 1086–1099 (2002).
11. K. Nishioka *et al*., *Mol. Cell* **9**, 1201–1213 (2002).
12. J. C. Rice *et al*., *Genes Dev.* **16**, 2225–2230 (2002).
13. S. Guyomarc'h, T. Vernoux, J. Traas, D. X. Zhou, M. Delarue, *J. Exp. Bot.* **55**, 673–684 (2004).
14. A. Perez-Riba, L. S. Itzhaki, *Curr. Opin. Struct. Biol.* **54**, 43–49 (2019).
15. L. Lu, X. Chen, S. Qian, X. Zhong, *Nat. Commun.* **9**, 630 (2018).
16. Y. Jacob *et al*., *Science* **343**, 1249–1253 (2014).
17. Y. Jacob *et al*., *Nat. Struct. Mol. Biol.* **16**, 763–768 (2009).
18. C. Raynaud *et al*., *Plant J.* **47**, 395–407 (2006).
19. Y. Jacob *et al*., *Nature* **466**, 987–991 (2010).
20. W. Feng *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **114**, 406–411 (2017).
21. H. Stroud *et al*., *PLOS Genet.* **8**, e1002808 (2012).
22. T. Suzuki *et al*., *Plant Cell Physiol.* **46**, 736–742 (2005).
23. J. M. Lucht *et al*., *Nat. Genet.* **30**, 311–314 (2002).
24. R. M. Williams, X. Zhang, *Prog. Biophys. Mol. Biol.* **163**, 109–119 (2021).
25. Q. Wang *et al*., *EMBO J.* **33**, 862–877 (2014).
26. J. A. Kamp, R. van Schendel, I. W. Dilweg, M. Tijsterman, *Nat. Commun.* **11**, 3615 (2020).
27. J. Dong *et al*., *Plant Cell* **33**, 961–979 (2021).
28. E. I. Campos *et al*., *Mol. Cell* **60**, 697–709 (2015).

ACKNOWLEDGMENTS

We thank X. Dong (Duke University) for providing us the seeds for the *brca2a* mutant (13F-1) and the recombination reporter line Col 1445; C. Bolick, E. Williams, and N. Guzzo (Yale) for help with plant growth and maintenance; K. Nelson for technical help with flow cytometry; members of the Yale Center for Genome Analysis; the Edinburgh Protein Production Facility (EPPF) for their support; and E. Campos (University of Toronto), Yale scientists W. Liu, J. Nijjer, and W. Yuan, and members of the Jacob and Couture labs for discussions, advice, and/or materials that contributed to this work. **Funding:** Supported by NIH grant R35GM128661 (Y.J.), grants from the Natural Science Engineering Research Council and the Canadian Institutes of Health Research (PJG-47391S) (J.-F.C.), an Ontario Graduate Scholarship and a University of Ottawa Excellence Scholarship (H.D.), a Yale University Brown Fellowship (B.M.), Fonds de Recherche du Québec–Nature et Technologies (FRQNT) grant 272565 (V.J.), and a Strategic Research Postdoctoral Fellowship from the University of Ottawa Heart Institute and the Strategic Research Endowed Funds (M.M.). Work in the Voigt lab was supported by Wellcome Trust grant 104175/Z/14/Z Sir Henry Dale Fellowship (P.V.) and UK Biotechnology and Biological Sciences Research Council grant BBS/E/B/000C0421. The Wellcome Centre for Cell Biology received core funding from Wellcome Trust grant 203149. The EPPF was supported by the Wellcome Trust through Multi-User Equipment grant 101527/Z/13/Z. B.H.R. and E.I.A. thank NSERC-RTI (RTI-2019-00009). **Author contributions:** Y.J. and J.-F.C. designed the research and supervised the study. Y.J., Y.-C.H., and J.D. designed and performed the initial experiments that identified the TPR domain of TSK/TONSL as an H3.1 reader. H.D. confirmed the H3.1-TSK interaction after doing the ITC assays, solving the crystal structure, and performing the in vitro structure-activity relationship study. Y.-C.H. performed the HR assays, and the analyses of the *tsk* and other DNA repair mutants, the H3.1S28A lines, and the H3.1 CRISPR mutants. J.D., H.D., and Y.-C.H. performed the in vitro binding assays, and A.A.-V. and J.D. contributed to the ITC assays. B.M. generated DNA repair mutants and contributed to their analysis. C.L. generated the H3.1S28A lines, and contributed with G.V. to in vivo experiments using these lines. A.P. performed the RNA-seq and DNA-seq analyses. G.T., V.J., and Y.-C.H. generated and validated the CRISPR mutants. M.M., A.R., B.H.R., and E.I.A. generated histone peptides. D.V. and P.V. designed and performed nucleosome pulldown experiments. J.S.B. collected structural data and generated a preliminary model. Y.J. and J.-F.C. wrote the manuscript with contributions from C.L., J.D., H.D., Y.-C.H., and P.V. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Sequencing data (DNA-seq and RNA-seq datasets) generated for this study are available at the Gene Expression Omnibus (GEO) under accession code GSE184738. The Protein Data Bank accession number for the TPR$_{TSK}$-H3.1$_{(1\text{-}45)}$ structure is 7T7T. All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm5320
Materials and Methods
Figs. S1 to S12
Tables S1 to S5
References (*29*–*64*)
MDAR Reproducibility Checklist

23 September 2021; accepted 27 January 2022
10.1126/science.abm5320

METALLOENZYMES

Recovery of particulate methane monooxygenase structure and activity in a lipid bilayer

Christopher W. Koo[1], Frank J. Tucci[1], Yuan He[1], Amy C. Rosenzweig[1,2]*

Bacterial methane oxidation using the enzyme particulate methane monooxygenase (pMMO) contributes to the removal of environmental methane, a potent greenhouse gas. Crystal structures determined using inactive, detergent-solubilized pMMO lack several conserved regions neighboring the proposed active site. We show that reconstituting pMMO in nanodiscs with lipids extracted from the native organism restores methane oxidation activity. Multiple nanodisc-embedded pMMO structures determined by cryo–electron microscopy to 2.14- to 2.46-angstrom resolution reveal the structure of pMMO in a lipid environment. The resulting model includes stabilizing lipids, regions of the PmoA and PmoC subunits not observed in prior structures, and a previously undetected copper-binding site in the PmoC subunit with an adjacent hydrophobic cavity. These structures provide a revised framework for understanding and engineering pMMO function.

Methane, a potent greenhouse gas, is a major contributor to the current climate crisis (*1*, *2*). Methane-oxidizing (methanotrophic) bacteria not only consume ~30 million metric tons of atmospheric methane per year (*3*) but also have the biotechnological potential to convert this cheap and abundant feedstock to fuels and value-added chemicals (*4*). Although methanotrophs have been engineered to produce a range of products, low yields and conversion efficiencies have precluded economic viability (*5*). For methane bioconversion to be transformative, the initial step—oxidation of methane to methanol—must be optimized, which requires molecular-level understanding of the main enzyme responsible, particulate methane monooxygenase (pMMO) (*6*, *7*).

pMMO is a membrane-bound, copper-dependent enzyme comprising three subunits, PmoA (β), PmoB (α), and PmoC (γ), arranged as a trimer of αβγ protomers. The crystal structures of detergent-solubilized pMMO from multiple methanotrophic species (*8*–*12*) have revealed the presence of three copper-binding sites. The ligands to two monocopper sites located in PmoB, the bis-His site and the $\mathrm{Cu_B}$ site, are not conserved, with the bis-His site only present in pMMOs of Gammaproteobacteria and the $\mathrm{Cu_B}$ site missing in pMMOs of Verrucomicrobia (*13*). By contrast, one aspartic acid and two histidine ligands to the third site, $\mathrm{Cu_C}$, located in PmoC, are strictly conserved. This observation, along with the saturated coordination geometry of the $\mathrm{Cu_B}$ site and the correlation of increased methane oxidation activity with copper occupancy of PmoC (*14*), suggests that the copper active site is located in PmoC.

However, it remains unclear whether the crystallographic $\mathrm{Cu_C}$ site is an appropriate active site model owing to several major caveats with the crystal structures. First, pMMO activity decreases upon solubilization in detergent (*11*, *15*), and purified samples exhibit zero methane oxidation activity (table S1), which means that the structures do not represent the active enzyme. Second, ~25 residues within PmoC are not observed in the electron density maps for any pMMO crystal structure. These residues, which correspond to the most highly conserved part of the PmoC sequence, are predicted to reside adjacent to the $\mathrm{Cu_C}$ site facing the interior of the pMMO trimer (*6*). The ambiguity in this region has precluded the identification of any potential cavities for methane and oxygen binding. It is likely that both of these limitations—the loss of activity and the disorder in PmoC—are attributable to the removal of pMMO from its native membrane environment. Disruption of the lipid bilayer followed by detergent solubilization and multiple purification steps may cause conformational changes and the separation of bound metal and/or lipid cofactors (*16*, *17*).

We recently demonstrated that reconstitution of detergent-solubilized pMMO into synthetic lipid bicelles (*11*) and nanodiscs (*14*) restores methane oxidation activity (table S1). To prepare enzymatically active samples for structure determination by cryo–electron microscopy (cryo-EM), *Methylococcus capsulatus* (Bath) pMMO was solubilized in *n*-dodecyl-β-D-maltoside (DDM) and embedded into nanodiscs using the membrane scaffold protein MSP1E3D1 and the commonly used lipids 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) and 1-palmitoyl-2-oleoylphosphatidylcholine (POPC) (*18*). To more closely mimic the cellular

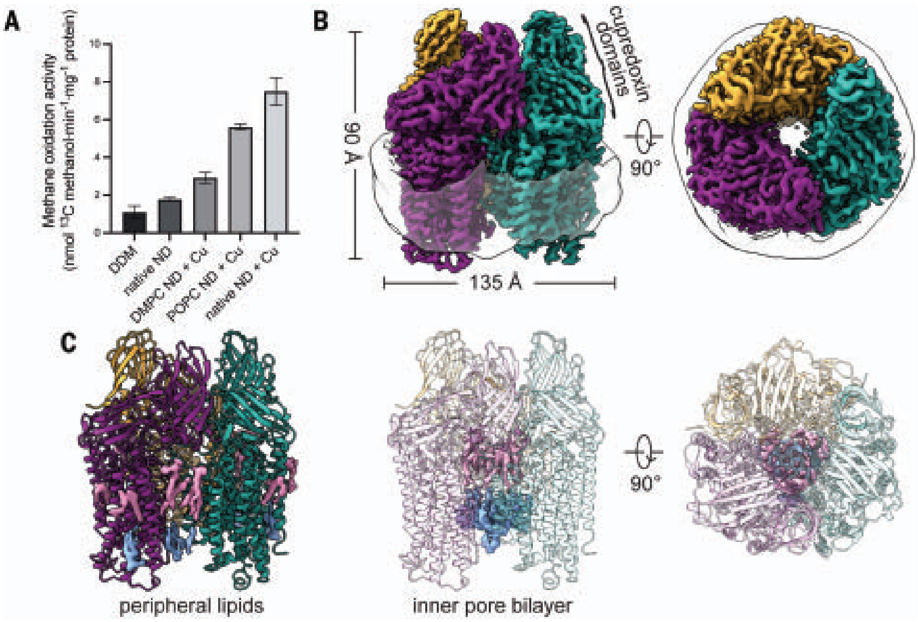


Fig. 1. Structural characterization of *M. capsulatus* (Bath) pMMO in a native lipid nanodisc. (**A**) Methane oxidation activity of cryo-EM samples. Addition of 3 equivalents of $\mathrm{CuSO_4}$ during nanodisc (ND) reconstitution (+ Cu) improved activity. Error bars represent standard deviations of $n \geq 3$ biological replicates, each measured in triplicate. (**B**) Cryo-EM map (dataset MC01) showing the trimer and the encircling nanodisc. Symmetrical αβγ protomers are colored in purple, teal, and gold, respectively. (**C**) Model of pMMO showing the map for lipids on the periphery and in the inner pore of the enzyme. The periplasmic and cytoplasmic leaflets are colored in pink and blue, respectively.

---

[1]Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA. [2]Department of Chemistry, Northwestern University, Evanston, IL 60208, USA.
*Corresponding author. Email: amyr@northwestern.edu

environment of pMMO, native lipids were extracted from *M. capsulatus* (Bath) (*19*, *20*) and used for nanodisc reconstitution. Nanodisc formation and pMMO incorporation were confirmed by negative-stain EM and cryo-EM (fig. S1).

As observed previously for *Methylocystis* species (sp.) strain (str.) Rockwell pMMO (*14*), the addition of copper during nanodisc reconstitution was necessary to recover methane oxidation activity. Nanodiscs formed with 3 molar equivalents of $CuSO_4$ and the native lipids exhibited the most activity, followed by POPC and DMPC nanodiscs (Fig. 1A and fig. S2). The copper content also increased slightly in the native and POPC nanodiscs (fig. S3). The looser packing afforded by the unsaturated bond in the acyl tails of POPC (*11*) and the native lipids (vide infra) may facilitate the loading of individual copper sites as well as access of the reductant used in the activity assays, duroquinol, to the active site. The observed turnover frequency of 0.012 s^{-1} represents a substantial improvement over the zero activity Cymal-5 samples used for previous structural studies and is comparable to *M. capsulatus* (Bath) membrane-bound pMMO activity measured with duroquinol (0.025 to 0.042 s^{-1}) (table S1). Notably, duroquinol is a synthetic analog of endogenous quinols (*21*) and is not the physiological reductant as has been claimed recently (*22*). In vivo electron delivery has been proposed to occur through ubiquinol reduced by a type 2 NADH:oxidoreductase and/or by coupling to methanol oxidation by methanol dehydrogenase (MDH) (*6*). The absence of these potential redox partners may be what precludes attaining whole-cell activity using isolated pMMO. Moreover, negative-stain images of *M. capsulatus* (Bath) membranes show dense packing of pMMO trimers (fig. S4), an environment that may contribute to increased activity in the membrane and that cannot be recapitulated in a single particle study.

Six cryo-EM maps of nanodisc-embedded pMMO from three different methanotrophs were obtained to 2.14- to 2.46-Å resolution (figs. S5 and S6 and table S2), providing the first structures of pMMO in a lipid environment. The highest-resolution map is that of enzymatically active *M. capsulatus* (Bath) pMMO in native lipid nanodiscs, resolved to 2.14 Å (MC01; table S2 and Fig. 1B), which is considerably higher resolution than that of the *M. capsulatus* (Bath) (2.8 Å) (*8*), the *Methylocystis* sp. str. Rockwell (2.6 Å) (*10*), and the *Methylotuvimicrobium alcaliphilum* combination (comb.) nova (nov.) 20Z (2.7 Å) (*11*) pMMO crystal structures. The overall architecture agrees with the crystal structures, consisting of a trimer of αβγ protomers within the nanodisc belt (Fig. 1B). Densities corresponding to solvent molecules and phospholipids (Fig. 1C and fig. S7) are clearly defined at this resolution.

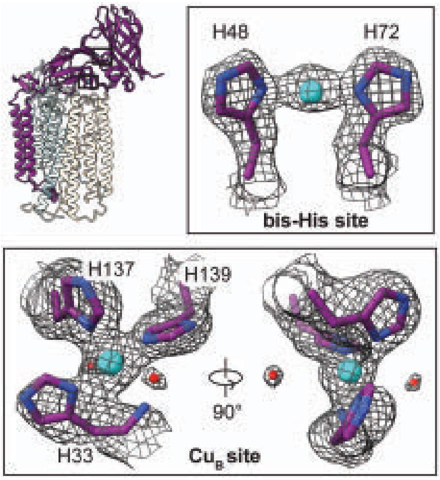


Fig. 2. Copper sites in the PmoB subunit of *M. capsulatus* (Bath) pMMO. The protomer model is shown with the PmoB subunit highlighted and the regions housing the copper sites boxed. Copper ions and water molecules are shown as cyan and red spheres, respectively. The cryo-EM map (MC01) is shown as a gray mesh.

The two periplasmic cupredoxin domains of PmoB extend out of the nanodisc (Fig. 1B) and exhibit the same folds as observed crystallographically, with two copper sites readily apparent in each cryo-EM map. The Cu_B site is coordinated by His^{33}, His^{137}, His^{139}, and the N terminus of PmoB in a square planar geometry (Fig. 2). Additionally, two water molecules, located 3.6 to 3.7 Å from the copper ion, stabilize the site through hydrogen-bonding interactions. Although axial water ligands were identified in crystallographic and spectroscopic studies of Cu_B (*10*, *23*–*25*), the observed water molecules are too distant for coordination. As in the crystal structures, the density for His^{33} and the N terminus is less clear, which suggests that it is somewhat flexible. The Cu_B site is unequivocally mononuclear (fig. S8), in contrast to a recent cryo-EM structure of *M. capsulatus* (Bath) pMMO in DDM (*26*) but consistent with recent spectroscopic (*23*, *24*, *27*) and computational (*25*) studies. The corresponding Cu_B sites in the cryo-EM structures of pMMO from *Methylocystis* sp. str. Rockwell and *M. alcaliphilum* comb. nov. 20Z exhibit the same mononuclear structures (RW01 and 20Z01; table S2 and fig. S8).

The bis-His site, coordinated by residues His^{48} and His^{72} (Fig. 2), is also present, consistent with the *M. capsulatus* (Bath) pMMO crystal structure (*8*). This site is occupied in the *M. alcaliphilum* comb. nov. 20Z pMMO cryo-EM map (fig. S8), despite being unoccupied in the crystal structure (*11*). The ligand corresponding to His^{48} is replaced with asparagine in *Methylocystis* sp. str. Rockwell pMMO, and the bis-His site is unoccupied (fig. S8). Whereas a recent cryo-EM study of *M. capsulatus* (Bath) pMMO in DDM suggested the presence of three additional copper ions in PmoB (*26*), no density at these specific locations is apparent in any of the maps of active pMMO in nanodiscs (fig. S9). There is no support in any of the structures for prior claims that a so-called Cu(I) sponge in PmoB mediates electron transfer (*26*, *28*).

In the transmembrane region, lipids coat the exterior of pMMO (Fig. 1C), with a total of 36 lipids modeled into the map as phosphatidylcholines along with 18 acyl chains, of which the head groups are not visible (fig. S7). A bilayer is observed in the interior of the pMMO trimer, showing that what appears to be a pore in the crystal structures is filled by phospholipids when embedded in a membrane-like environment (Fig. 1C). The longest acyl chain could be resolved to 12 carbons, with the density becoming less clear as the tails extend toward the middle of the bilayer. To investigate the possible identities of these native phospholipids, both *M. capsulatus* (Bath) whole cells and the native lipid pMMO-nanodisc complex were subjected to lipidomics analysis by liquid chromatography–tandem mass spectrometry (LC-MS/MS). A mixture of phosphatidylethanolamine (PE), phosphatidylcholine (PC), phosphatidylglycerol (PG), and cardiolipin (CL) was identified (fig. S10). For comparison, the structure of *M. capsulatus* (Bath) pMMO in a POPC nanodisc was determined to 2.26-Å resolution (MC02; table S2). The lipid densities appear very similar between the two maps, which suggests that most of the observed lipids are PC lipids (fig. S11), although other native lipids may remain associated with pMMO through the nanodisc reconstitution process. One exterior lipid density exhibits an unusual shape, which suggests that it may have a different identity, such as a native quinone that remains bound during purification and nanodisc reconstitution (fig. S11). This density occupies the position filled by an unidentified helix in the *Methylocystis* sp. str. Rockwell pMMO crystal structure (*10*) and in the cryo-EM structure (RW01).

Several regions of *M. capsulatus* (Bath) pMMO that were not observed in the crystal structures are ordered in the cryo-EM maps. PmoA residues 192 to 212 are stabilized by phospholipids on the interior and exterior of the enzyme (Fig. 3A). This part of PmoA extends into PmoC, with PmoA residue Arg^{206} forming a hydrogen bond with PmoC residue Glu^{238} (Fig. 3B), a strictly conserved residue not observed in the crystal structures (Fig. 3, C and D). There is no evidence for a tricopper D site proposed to be depleted from the inactive pMMO (*29*). Although two of the three copper ions in this putative site were recently modeled in the cryo-EM structure of *M. capsulatus* (Bath) pMMO in DDM (*26*), our high-resolution maps instead show Glu^{154} and a water molecule in the region (fig. S12).

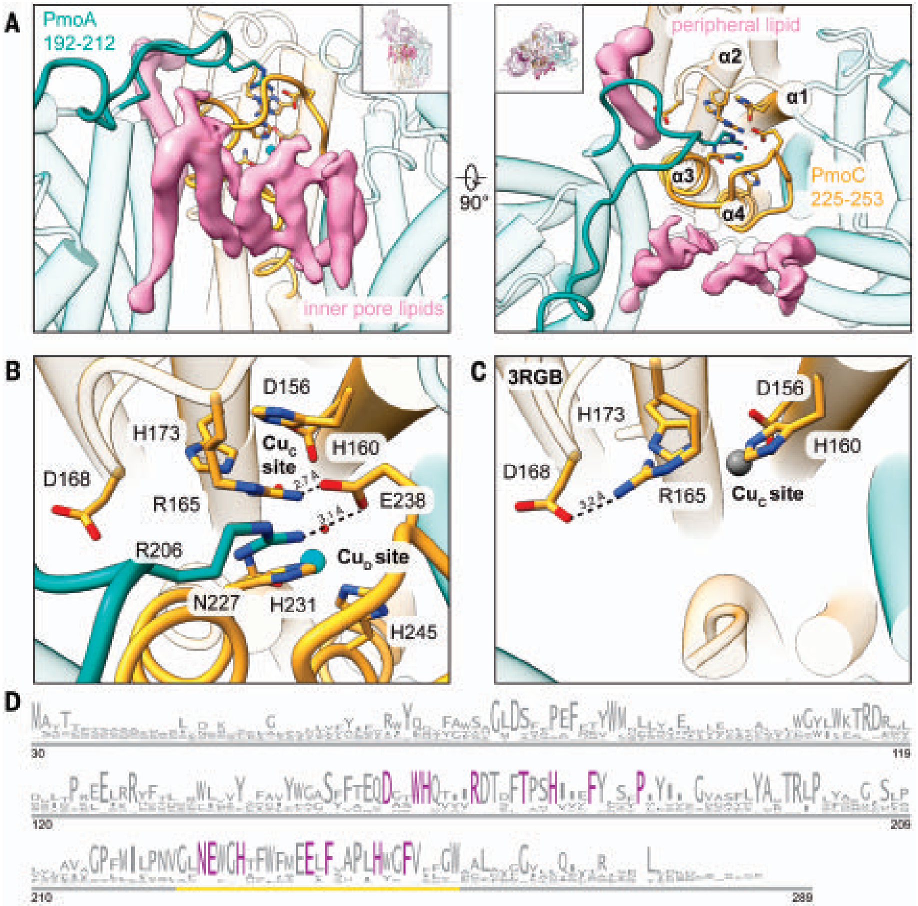


Fig. 3. Regions of *M. capsulatus* (Bath) pMMO newly observed in the cryo-EM structures. (**A**) Newly modeled residues in PmoA (192 to 212) (left) and PmoC (225 to 253) (right) are highlighted in teal and gold, respectively. Cryo-EM densities corresponding to lipids that interact with these regions are shown in pink. (**B**) Magnified view of the completed four-helix bundle from the cryo-EM structure (MC01). Key residues are labeled, and the hydrogen-bonding network involving Arg^{206} (PmoA), Glu^{238} (PmoC), and Arg^{165} (PmoC) is marked with dashed lines. The Cu$_D$ site copper ion is shown as a cyan sphere. (**C**) Magnified view of the same region shown in (B) in the crystal structure (Protein Data Bank ID: 3RGB) (*9*), with the hydrogen bond between Asp^{168} and Arg^{165} shown as a dashed line. The zinc ion occupying the Cu$_C$ site is shown as a gray sphere. (**D**) Sequence logo for the PmoC subunit (PF04896). Only sequences found in pMMO or ammonia monooxygenase operons are included. Residues conserved at 100% are colored in purple. The gold line demarcates the newly stabilized region in the cryo-EM structures. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

This water molecule forms hydrogen bonds with putative tricopper ligands Glu^{100} from PmoA and Glu^{154} from PmoC, which in turn is hydrogen bonded to Asn^{103} from PmoA, also claimed to be a ligand (*26*). The other hypothetical ligands, His^{38} and Met^{42}, do not interact with any additional densities that could be attributed to copper ions (fig. S12). Moreover, a suggested hydrophobic pocket for substrate binding involving Trp^{48}, Phe^{50}, Trp^{51}, and Trp^{54} (*30*) is occupied by Asp^{47} from PmoA.

The highly conserved region of PmoC (Fig. 3D) that was unmodeled in the crystal structures (Fig. 3C and fig. S13) and the recent cryo-EM structure of pMMO in DDM (*26*) is observed in all the cryo-EM maps (figs. S13 and S14). This region corresponds to residues 225 to 253 in *M. capsulatus* (Bath) pMMO. Hydrophobic residues from this sequence interact with phospholipid tails in the trimer interior, including close interactions between phospholipids and previously unmodeled residues (fig. S14). These interactions stabilize a four-helix bundle comprising residues 123 to 163 (α1), 170 to 199 (α2), 220 to 233 (α3), and 244 to 270 (α4) (Fig. 3A). The Cu$_C$ ligands, Asp^{156}, His^{160}, and His^{173}, derive from helices α1 and α2. This structure is observed in each of the *M. capsulatus* (Bath) pMMO maps, including the POPC nanodisc (MC02; table S2), which suggests that the stabilizing effect is imparted by the bilayer rather than a particular native lipid. The nanodisc and lipid environment may restore activity to pMMO (Fig. 1A) by stabilizing catalytically important features of this highly conserved region.

The environment of the crystallographically modeled Cu$_C$ site is drastically altered in the cryo-EM structure. Newly observed helices α3 and α4 house three strictly conserved residues, Asn^{227}, His^{231}, and His^{245}, which extend toward the crystallographic Cu$_C$ ligands Asp^{156}, His^{160}, and His^{173} (Fig. 3B). Residue Arg^{165}, which was hydrogen-bonded to Asp^{168} in the crystal structure (Fig. 3C), is now positioned near the Cu$_C$ site via a hydrogen-bonding network to Glu^{238} and PmoA Arg^{206} (Fig. 3B and Fig. 4A). Arginine residues in hydrophobic environments play a variety of functional roles in enzymes, including regulating redox potentials and substrate binding (*31*, *32*). Unexpectedly, Asn^{227}, His^{231}, and His^{245} are connected to a strong density that distinctly resembles a metal ion (Fig. 4B). This metal ion is ligated by the side-chain ε nitrogens of the two histidine residues (Cu-N; 2.0 Å) and the side chain of Asn^{227} (2.35 Å) in a trigonal planar geometry (Fig. 4B). The metal ion density is more persistent than the surrounding protein density at higher thresholds and is present in the cryo-EM maps of three independent samples of *M. capsulatus* (Bath) pMMO in both native (table S2; MC01 and MC04) and POPC (table S2; MC02) nanodiscs (Fig. 4B and fig. S15). Given that copper restores activity to metal-depleted membranes (*10*) and to nanodisc samples (*14*), it is likely that the density corresponds to a previously unknown copper site, denoted here as Cu$_D$ [not to be confused with the proposed D site in PmoA (*26*)]. The Glu^{228} carboxylate group is within hydrogen-bonding distance of the Asn^{227} side chain, consistent with hydrogen bonding via its side-chain amide group and coordination of copper by its side-chain oxygen atom (fig. S16). Copper coordination by the side-chain oxygen of asparagine or glutamine is unusual but not unprecedented (*33*–*36*). Recent X-band electron paramagnetic resonance (EPR) and electron nuclear double resonance (ENDOR) spectroscopic studies of *M. capsulatus* (Bath) pMMO in POPC nanodiscs indicate the presence of two histidine ligands to the copper ion spectroscopically assigned as Cu$_C$ (*27*), a finding that would be consistent with either the structural Cu$_C$ or Cu$_D$ site.

In contrast to the Cu$_D$ site, the Cu$_C$ site is only apparent in one map of *M. capsulatus* (Bath) pMMO (table S2 and fig. S15; MC03), with strong density connected to that of the three Cu$_C$ ligands, Asp^{156} (2.7 Å), His^{160} (2.1 Å), and His^{173} (2.2 Å). The presence of

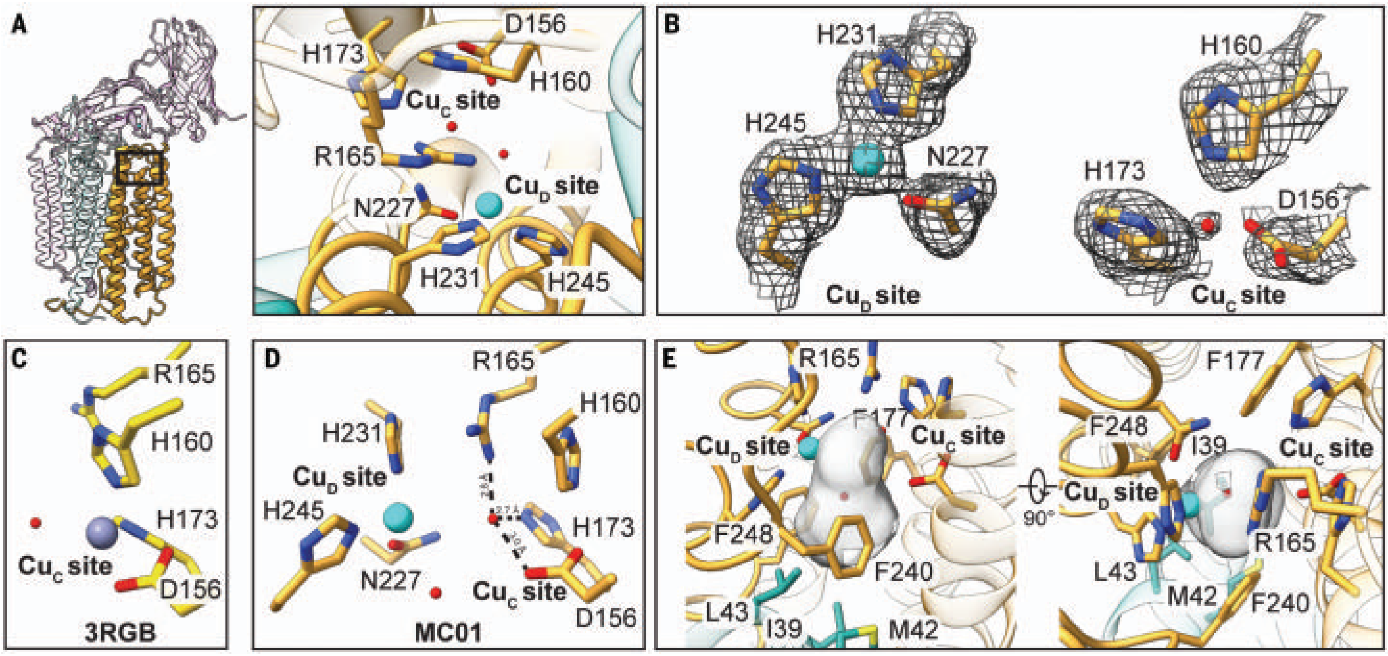


Fig. 4. Active site architecture in the PmoC subunit of *M. capsulatus* (Bath) pMMO. (**A**) The protomer is shown with the PmoC subunit highlighted and the region housing the metal-binding sites boxed (left) and as a magnified top-down view (right). (**B**) The map (MC01) at the Cu_D and Cu_C sites. (**C**) The corresponding model for the crystal structure (Protein Data Bank ID: 3RGB) (*9*) in which the Cu_C site is occupied by zinc. The model lacks residues 225 to 253, which include the Cu_D site residues. (**D**) Overall model of Cu_C and Cu_D sites, with hydrogen bonds involving the water molecule in the Cu_C site shown as dashed lines. (**E**) Cavity located between and below the two sites shown in gray, generated using CASTp (*42*). Residues lining the cavity are labeled.

Cu_C correlates with increased disorder in the loop (residues 233 to 240) connecting newly visible helices α3 and α4 in PmoC. Although density for the Cu_D ligands is still observed, there is no density attributable to a metal ion (fig. S15). Similarly, the *M. alcaliphilum* comb. nov. 20Z pMMO map (table S2; 20Z01) exhibits a clear Cu_C site (fig. S17A), but the density for the region corresponding to *M. capsulatus* (Bath) PmoC residues 233 to 240 (residues 206 to 213 in *M. alcaliphilum* comb. nov. 20Z PmoC) is poorly defined (fig. S14A) and completely lacks side-chain density for Arg^{165} (Arg^{137} in *M. alcaliphilum* comb. nov. 20Z PmoC; fig. S17A). There is no obvious density for Cu_D, and nanodisc samples of *M. alcaliphilum* comb. nov. 20Z pMMO exhibit no methane oxidation activity (table S1). The Cu_C site is again observed in the *Methylocystis* sp. str. Rockwell pMMO map, and the region containing the Cu_D ligands (residues 200 to 221), although ordered (fig. S14A), lacks side-chain density (fig. S17B). Upon further processing of this map using DeepEMhancer (*37*), the Cu_D site ligands were resolved (fig. S17B). In all six maps, the density at the Cu_C or Cu_D sites is not as strong as that observed at Cu_B, which is consistent with the full Cu_B occupancy observed by crystallography and native top-down mass spectrometry (nTDMS) (*14*).

In two of the three *M. capsulatus* (Bath) pMMO maps that show density for Cu_D (table S2; MC01 and MC02), there is no density between the Cu_C ligands. By contrast, the crystal structure contains zinc derived from the crystallization buffer at the Cu_C site (Fig. 4C), and crystal structures of pMMO isolated from *Methylocystis* sp. str Rockwell contain copper in this site (*10*). In the cryo-EM maps lacking density for Cu_C, residue His^{160} adopts a slightly different conformation, presumably because it is no longer stabilized by copper coordination. Instead of Cu_C, a spherical density ~2 Å from the location of Cu_C is observed (Fig. 4B). This density was modeled as a water molecule because it is within hydrogen-bonding—rather than coordinating—distance of Cu_C ligands Asp^{156} and His^{173} as well as Arg^{165} (Fig. 4D). This water molecule is not present in the third map (table S2; MC04). Instead, the density corresponding to Cu_C ligands His^{160} and His^{173} is weakly connected (fig. S18), offering the possibility that the Cu_C and Cu_D sites, separated by ~5.7 Å, might be occupied simultaneously.

Notably, the cryo-EM model of active *M. capsulatus* (Bath) pMMO reveals a hydrophobic cavity adjacent to the Cu_D site (Fig. 4E) extending away from the periplasmic side of pMMO. This cavity is lined by PmoA residues Ile^{39}, Met^{42}, and Leu^{43} and PmoC residues Phe^{177}, Phe^{240}, and Phe^{248}. These three PmoC residues are invariant (Fig. 3D), and the latter two derive from the newly observed region of PmoC. This cavity houses the aforementioned water molecule and abuts the location of the Cu_D ligands (Fig. 4D). Its existence supports assignment of the pMMO active site as one or both of the two copper sites, Cu_C and Cu_D.

All four *M. capsulatus* (Bath) pMMO samples had similar activities and copper contents (Fig. 1A and fig. S3), precluding correlation of the Cu_C and Cu_D occupancies with methane oxidation. Treating the nanodiscs with excess copper after reconstitution did not increase the apparent occupancy of either site and was found to inhibit activity, consistent with previous studies showing the inhibitory effect of excess copper (*10*). Although we assigned Cu_D as copper on the basis of extensive data linking copper addition to restored activity (*10*, *14*), cryo-EM data, unlike x-ray diffraction data (*38*), do not allow for unambiguous metal identification. Other possibilities include iron and zinc, but iron does not restore activity to metal-depleted membranes (*39*), and zinc inhibits activity (*10*). Additionally, only heme iron, attributed to cytochrome impurities, has been detected spectroscopically in pMMO (*40*).

To further probe the identity of the metal ion at Cu_D and its role in methane oxidation, metals were removed from *M. capsulatus* (Bath) membranes using potassium cyanide (*10*, *41*), and the sample was split into two batches, one of which was treated with 10 equivalents of $CuSO_4$. pMMO was solubilized from both batches, reconstituted into nanodiscs, and investigated by cryo-EM. A cryo-EM map of the metal-depleted sample, which contained 0.1 equivalents of Cu (fig. S19), was resolved to 3.65 Å (fig. S20A and table S2; MC05). Some density is still observed at Cu_B, but the occupancy of the bis-His site is reduced substantially (fig. S21A). Whereas the PmoB and PmoA subunits remain intact, the PmoC subunit is completely disrupted (fig. S21A), and the sample exhibits no methane oxidation activity (fig. S22). There is no density for residues 54 to 97, which make up the first helix; residues 160 to 178, including the Cu_C site; and residues 221 to 246, including the Cu_D site. The disordered regions are bracketed by coordinating histidines (Cu_C His^{160} and Cu_D His^{245}), which indicates that PmoC requires copper for structural stability. In addition, all of the stabilized lipids are disrupted in this structure.

The cryo-EM map of the metal-depleted, copper-reloaded sample, which contained ~0.6 equivalents of copper (fig. S19), was resolved to 3.39 Å (fig. S20B and table S2; MC06). Despite the lower resolution, lipids on the inner pore and periphery of pMMO are clearly visible, and the backbone could be traced through the entire PmoC subunit (fig. S21B). Thus, the destabilizing effects of copper removal on structure can be reversed upon copper addition. The Cu_B and bis-His sites are more fully occupied (fig. S21B) as well. Although residues at the Cu_D site are less well defined, there is density near His^{245} and Asn^{227} that could correspond to a metal ion (fig. S21C), supporting the assignment of Cu_D as a copper-binding site. The Cu_C site residues are well defined and oriented toward one another, but the site appears unoccupied (fig. S21C). Notably, this sample also exhibits methane oxidation activity (fig. S22) on par with that observed in pMMO reconstituted into native nanodiscs with no additional copper added (Fig. 1A) and consistent with previous studies showing that ~50% of activity can be regained after cyanide treatment and copper reloading (*10*). This result links reloading of Cu_D to the recovery of some methane oxidation activity.

These first structures of active pMMO, obtained by embedding the enzyme in a native lipid bilayer, provide critical insight into pMMO structure and function. The combined results indicate that an intact PmoC scaffold, including Cu_C and Cu_D ligands supported by a hydrogen-bonding network and interior lipid bilayer, is associated with enzymatic activity. The presence of Cu_D in most of the *M. capsulatus* (Bath) cryo-EM structures suggests a role in activity, but the occupancy of the Cu_C and Cu_D sites in vivo remains to be determined. The structure of the highly conserved region of PmoC—along with the unexpected discovery of the Cu_D site and adjacent hydrophobic cavity—provide a much more complete picture of the pMMO active site architecture. This substantially revised view of pMMO, obtained after >15 years of crystallographic characterization, underscores the importance of studying membrane proteins in their native environments and the potential of high-resolution cryo-EM combined with membrane mimetic technology.

REFERENCES AND NOTES

1. B. Barrett, J. W. Charles, J. L. Temte, *Prev. Med.* **70**, 69–75 (2015).
2. X. L. Chai, D. J. Tonjes, D. Mahajan, *Prog. Energy Combust. Sci.* **56**, 33–70 (2016).
3. C. L. Curry, *Global Biogeochem. Cycles* **21**, GB4012 (2007).
4. I. Y. Hwang *et al*., *Appl. Microbiol. Biotechnol.* **102**, 3071–3080 (2018).
5. A. D. Nguyen, E. Y. Lee, *Trends Biotechnol.* **39**, 381–396 (2021).
6. C. W. Koo, A. C. Rosenzweig, *Chem. Soc. Rev.* **50**, 3424–3436 (2021).
7. M. O. Ross, A. C. Rosenzweig, *J. Biol. Inorg. Chem.* **22**, 307–319 (2017).
8. R. L. Lieberman, A. C. Rosenzweig, *Nature* **434**, 177–182 (2005).
9. S. M. Smith *et al*., *Biochemistry* **50**, 10231–10240 (2011).
10. S. Sirajuddin *et al*., *J. Biol. Chem.* **289**, 21782–21794 (2014).
11. S. Y. Ro *et al*., *J. Biol. Chem.* **293**, 10457–10465 (2018).
12. A. S. Hakemian *et al*., *Biochemistry* **47**, 6793–6801 (2008).
13. C. W. Koo, A. C. Rosenzweig, "Particulate methane monooxygenase and the PmoD protein" in *Encyclopedia of Inorganic and Bioinorganic Chemistry*, R. A. Scott, Ed. (Wiley, 2020).
14. S. Y. Ro *et al*., *Nat. Commun.* **10**, 2675 (2019).
15. S. Sirajuddin, A. C. Rosenzweig, *Biochemistry* **54**, 2283–2294 (2015).
16. R. M. Garavito, S. Ferguson-Miller, *J. Biol. Chem.* **276**, 32403–32406 (2001).
17. A. Anishkin, S. H. Loukin, J. Teng, C. Kung, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 7898–7905 (2014).
18. T. H. Bayburt, S. G. Sligar, *FEBS Lett.* **584**, 1721–1727 (2010).
19. E. G. Bligh, W. J. Dyer, *Can. J. Biochem. Physiol.* **37**, 911–917 (1959).
20. J. Folch, M. Lees, G. H. Sloane Stanley, *J. Biol. Chem.* **226**, 497–509 (1957).
21. S. A. Cook, A. K. Shiemke, *Arch. Biochem. Biophys.* **398**, 32–40 (2002).
22. W. Peng, X. Qu, S. Shaik, B. Wang, *Nat. Catal.* **4**, 266–273 (2021).
23. M. O. Ross *et al*., *Science* **364**, 566–570 (2019).
24. G. E. Cutsail 3rd, M. O. Ross, A. C. Rosenzweig, S. DeBeer, *Chem. Sci.* **12**, 6194–6209 (2021).
25. L. Cao, O. Caldararu, A. C. Rosenzweig, U. Ryde, *Angew. Chem. Int. Ed.* **57**, 162–166 (2018).
26. W. H. Chang *et al*., *J. Am. Chem. Soc.* **143**, 9922–9932 (2021).
27. R. J. Jodts *et al*., *J. Am. Chem. Soc.* **143**, 15358–15368 (2021).
28. Y. J. Lu *et al*., *J. Inorg. Biochem.* **196**, 110691 (2019).
29. S. I. Chan *et al*., *Angew. Chem. Int. Ed.* **46**, 1992–1994 (2007).
30. K. Y. Ng, L. C. Tu, Y. S. Wang, S. I. Chan, S. S. Yu, *ChemBioChem* **9**, 1116–1123 (2008).
31. M. J. Harms, J. L. Schlessman, G. R. Sue, B. García-Moreno E., *Proc. Natl. Acad. Sci. U.S.A.* **108**, 18954–18959 (2011).
32. C. W. Reid, N. T. Blackburn, A. J. Clarke, *Biochemistry* **45**, 2129–2138 (2006).
33. N. Vita *et al*., *Sci. Rep.* **6**, 39065 (2016).
34. C. Dennison, M. D. Harrison, *J. Am. Chem. Soc.* **126**, 2481–2489 (2004).
35. P. J. Hart *et al*., *Protein Sci.* **5**, 2175–2183 (1996).
36. Y. Xie *et al*., *J. Biochem.* **137**, 455–461 (2005).
37. R. Sanchez-Garcia *et al*., *Commun. Biol.* **4**, 874 (2021).
38. S. E. J. Bowman, J. Bridwell-Rabb, C. L. Drennan, *Acc. Chem. Res.* **49**, 695–702 (2016).
39. R. Balasubramanian *et al*., *Nature* **465**, 115–119 (2010).
40. R. L. Lieberman *et al*., *Inorg. Chem.* **45**, 8372–8381 (2006).
41. S. M. Smith, R. Balasubramanian, A. C. Rosenzweig, *Methods Enzymol.* **495**, 195–210 (2011).
42. W. Tian, C. Chen, X. Lei, J. Zhao, J. Liang, *Nucleic Acids Res.* **46**, W363–W367 (2018).

ACKNOWLEDGMENTS

We thank R. Abdella for advice on data processing, S. Ro for advice on methanotroph handling, J. Pattie for computer support, and J. Remis and R. Purohit for guidance on sample preparation and data collection. Molecular graphics and analyses were performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health grant R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases. **Funding:** This work was supported by National Institutes of Health grants R35GM118035 (A.C.R.), T32GM008382 (C.W.K.), T32GM105538 (F.J.T.), and R01GM135651 (Y.H.). This work used resources of the Northwestern University Structural Biology Facility, which is generously supported by the NCI CCSG P30 CA060553 grant awarded to the Robert H. Lurie Comprehensive Cancer Center. We acknowledge the use of the Ametek K3 direct electron detector, which was generously provided by R. A. Lamb (HHMI investigator). A portion of this research was supported by NIH grant U24GM129547 and performed at the PNCC at OHSU and accessed through EMSL (grid.436923.9), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research. Some of this work was performed at the Stanford-SLAC Cryo-EM Center (S2C2) supported by the NIH Common Fund Transformative High-Resolution Cryoelectron Microscopy program (U24 GM129541). Some of this work was performed at the National Center for CryoEM Access and Training (NCCAT) and the Simons Electron Microscopy Center located at the New York Structural Biology Center, supported by the NIH Common Fund Transformative High-Resolution Cryoelectron Microscopy program (U24 GM129539) and by grants from the Simons Foundation (SF349247) and the NY State Assembly. **Author contributions:** Conceptualization: C.W.K., F.J.T., and A.C.R. Methodology: C.W.K., F.J.T., Y.H., and A.C.R. Investigation: C.W.K. and F.J.T. Visualization: C.W.K. and F.J.T. Funding acquisition: A.C.R. Project administration: A.C.R. Supervision: A.C.R. and Y.H. Writing – original draft: C.W.K., F.J.T., and A.C.R. Writing – review and editing: C.W.K., F.J.T., Y.H., and A.C.R. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The models of pMMO from *M. capsulatus* (Bath), pMMO from *M. alcaliphilum* comb. nov. 20Z, and pMMO from *Methylocystis* sp. str. Rockwell have been deposited in the Protein Data Bank with accession codes 7S4H (MC01), 7S4I (MC02), 7S4J (MC03), 7S4K (MC04), 7S4L (20Z01), 7S4M (RW01), 7T4O (MC05), and 7T4P (MC06). The corresponding cryo-EM maps are available at the Electron Microscopy Data Bank (www.ebi.ac.uk/emdb/). All other data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm3282
Materials and Methods
Figs. S1 to S22
Tables S1 and S2
References (*43–60*)

10 September 2021; resubmitted 13 December 2021
Accepted 16 February 2022
10.1126/science.abm3282

WILDFIRES

Wildfire smoke destroys stratospheric ozone

Peter Bernath[1,2,3]*, **Chris Boone**[2], **Jeff Crouse**[2]

Large wildfires inject smoke and biomass-burning products into the mid-latitude stratosphere, where they destroy ozone, which protects us from ultraviolet radiation. The infrared spectrometer on the Atmospheric Chemistry Experiment satellite measured the spectra of smoke particles from the "Black Summer" fires in Australia in late 2019 and early 2020, revealing that they contain oxygenated organic functional groups and water adsorption on the surfaces. These injected smoke particles have produced unexpected and extreme perturbations in stratospheric gases beyond any seen in the previous 15 years of measurements, including increases in formaldehyde, chlorine nitrate, chlorine monoxide, and hypochlorous acid and decreases in ozone, nitrogen dioxide, and hydrochloric acid. These perturbations in stratospheric composition have the potential to affect ozone chemistry in unexpected ways.

The occurrence and severity of wildfires have been increasing as a result of climate change (*1*). Recently, there have been several extreme wildfires, including the "Black Saturday" bushfires in Australia in 2009 (*2*), wildfires in the Pacific Northwest of North America in 2017 (*3*), and more severe "Black Summer" bushfires in Australia in 2019 and 2020 (*4*). Severe fires can create pyrocumulonimbus clouds (pyroCb) (*5*) that rapidly transport fire products into the upper troposphere and sometimes even into the lower stratosphere (*6*). PyroCbs can affect climate by heating the stratosphere and cooling Earth's surface (*7*); stratospheric chemistry may also be affected (*2*), and it is likely that climate change will also increase the occurrence of pyroCbs (*8*).

Extreme pyroCbs have been compared to moderate volcanic eruptions (*9*) and "nuclear winter" (*4*) because of the increase in stratospheric aerosols from smoke. Similar to large volcanic eruptions, pyroCbs can decrease stratospheric ozone (O_3), which shields us from damaging ultraviolet radiation (*10*). Atmospheric Chemistry Experiment (ACE) satellite (*11*) observations point to strong perturbations of stratospheric chemistry by pyroCbs that result in stratospheric ozone depletion at mid-latitudes. The Montreal Protocol on Substances that Deplete the Ozone Layer (*12*) has been successful in reducing the atmospheric abundances of chlorine- and bromine-containing molecules that destroy stratospheric ozone and are responsible for the Antarctic ozone hole (*13*, *14*). The increasing frequency of major wildfires (*1*, *8*), however, has the potential to delay the recovery of stratospheric ozone, which is currently predicted to return to 1980 levels around 2052–2060 (*15*).

The main instrument on the ACE satellite is a high-resolution infrared Fourier transform spectrometer (FTS), which has been making solar occultation observations of the atmosphere in a limb geometry since 2004 (*11*). During sunrise and sunset, the ACE-FTS records a sequence of atmospheric absorption spectra using the Sun as a light source. Note that ACE observations occur at twilight, and abundances of photochemically active molecules such as ClO and HOCl depend on the time at which the measurements are made. These infrared spectra are then analyzed on the ground (*16*) to provide altitude profiles of volume mixing ratios (VMRs) and concentrations for more than 44 molecules (*17*) as well as infrared spectra of aerosols and diffuse clouds. ACE also has a visible imager (at 0.525 μm) and a near-infrared imager (at 1.02 μm), both of which provide atmospheric extinction due to clouds and aerosols.

The ACE-FTS has recorded spectra of the material injected into the stratosphere by pyroCbs during the extreme Australian bushfires in late December 2019 to early January 2020 (*6*). Spectra of stratospheric smoke particles were obtained by removing absorption features as a result of gaseous molecules; a sample aerosol infrared spectrum is shown in Fig. 1. These characteristic spectra show strong absorption features due to C=O, CH, and OH stretching modes consistent with carboxylic acid [–C(O)OH] groups bound to carbon. There is also evidence of adsorbed water from the hydrogen-bonded OH stretching and librational modes. No features

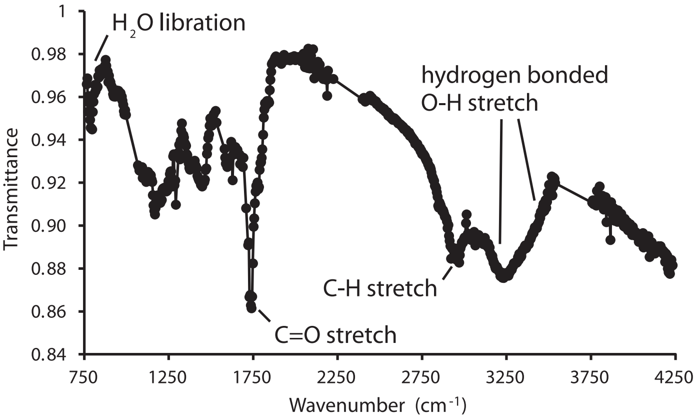


Fig. 1. The aerosol spectrum for tangent height 16.6 km from occultation ss88361, measured 6 January 2020 at latitude 58.8°S. In "ss8836," "ss" stands for sunset and 88361 is the number of orbits since launch, a unique identifier for the measurement. Spectral features associated with selected functional groups are indicated.

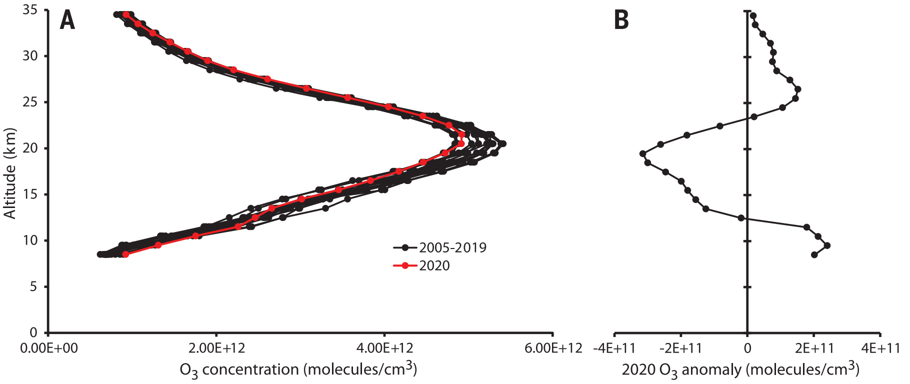


Fig. 2. ACE-FTS ozone concentration profiles. (**A**) The average O_3 concentration profiles for the latitude range 45° to 60°S for ACE-FTS occultations in July. The profiles for all years before 2020 measured by the ACE-FTS are shown in black and the average profile for 2020 is shown in red. (**B**) The anomaly profile for 2020 O_3 concentration is the 2020 profile in (A) minus the average profile from all other years.

---

[1]Department of Chemistry and Biochemistry, Old Dominion University, Norfolk, VA, USA. [2]Department of Chemistry, University of Waterloo, Waterloo, ON, Canada. [3]Department of Physics, Old Dominion University, Norfolk, VA, USA.
***Corresponding author. Email: pbernath@odu.edu**

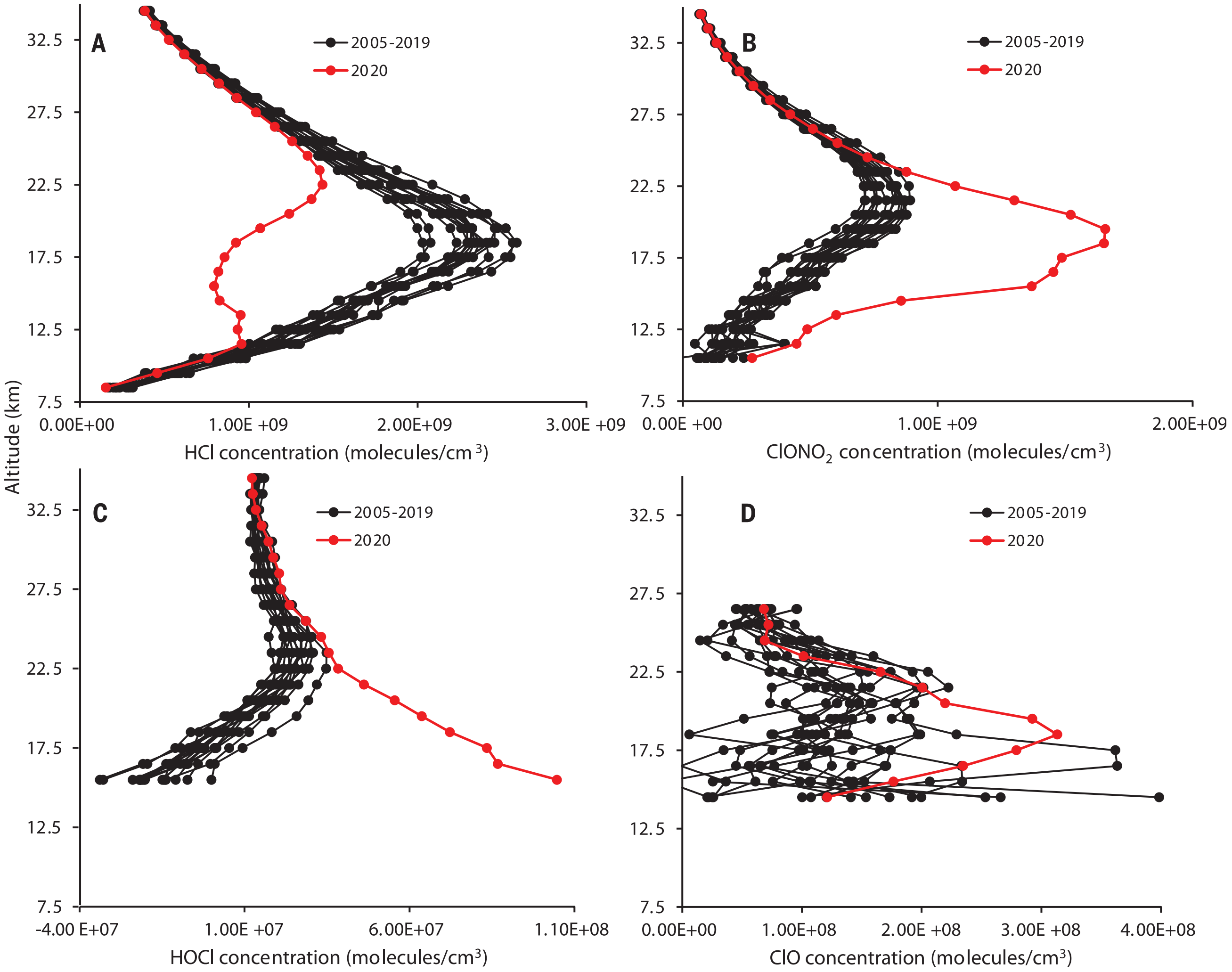


Fig. 3. Concentrations of chlorine-containing molecules for the latitude range 45° to 60°S for ACE-FTS occultations in late May and early June. Profiles for all years before 2020 measured by the ACE-FTS are shown in black and the average profile for 2020 is in red. (**A**) HCl. (**B**) $ClONO_2$. (**C**) HOCl. (**D**) ClO.

are assignable to adsorbed sulfuric acid (*7*). Surface reactions on this acidic, hydrated soot have not yet been studied in the laboratory and are hence unknown. However, here we describe how they are capable of perturbing several species expected to influence ozone, such as nitrogen dioxide and hypochlorous acid. ACE observed enhanced aerosols with the ACE imagers in the Southern Hemisphere stratosphere from January to November 2020, primarily because of the fires (fig. S1).

Average ACE ozone concentrations at midlatitudes (45° to 60°S) were found to increase in the 8- to 20-km range in January and March 2020 (fig. S2) compared with the average of all other years observed by ACE because of the injection of many organic molecules into the upper troposphere and lower stratosphere. The usual tropospheric oxidation chemistry responsible for ozone pollution (*18*) is therefore the cause of this early ozone increase after the pyroCb events. The lifetime of these organic molecules is relatively short, so this enhancement does not persist much beyond March 2020. Starting in April 2020, midlatitude ozone levels began to decline in the lower stratosphere compared with the average of all other years and remained low through to December 2020 (fig. S3). Fig. 2 shows the ozone profile for July 2020 for 45° to 60°S compared with the average of all other years. Similar ozone declines were observed from 30° to 45°S in the 16- to 26-km altitude range, but no significant declines were observed (5 to 35 km) over Antarctica (prior to Southern Hemisphere winter) or the tropics compared with previous years.

Some years had O_3 levels near 18 km during certain months that were lower or comparable to those observed in 2020: for example, 2005 and 2008 in July (Fig. 2), 2008 in May and June, and 2006 in October (fig. S3). However, no other year exhibited persistently low O_3 levels (compared with other years measured by ACE) over such an extended period of time. Later in the year (October and December; fig. S3), average O_3 near 18 km for 2020 in this latitude range was well below the typical range of variability established by measurements from previous years.

The stratospheric chemistry of ozone is well known to involve HO_x radicals (H, OH, and HO_2), halogen chemistry (Cl and Br), and reactive nitrogen species (*18*, *19*). Trace gases and soot from pyroCbs perturb each of these reaction families.

Unexpected and notable changes were observed in a range of lower stratospheric, chlorine-containing species from March to August 2020 at 45° to 60°S. ACE-FTS midlatitude observations show a decrease in hydrochloric acid (HCl) abundance but increases in chlorine nitrate ($ClONO_2$), hypochlorous acid (HOCl), and chlorine monoxide (ClO).

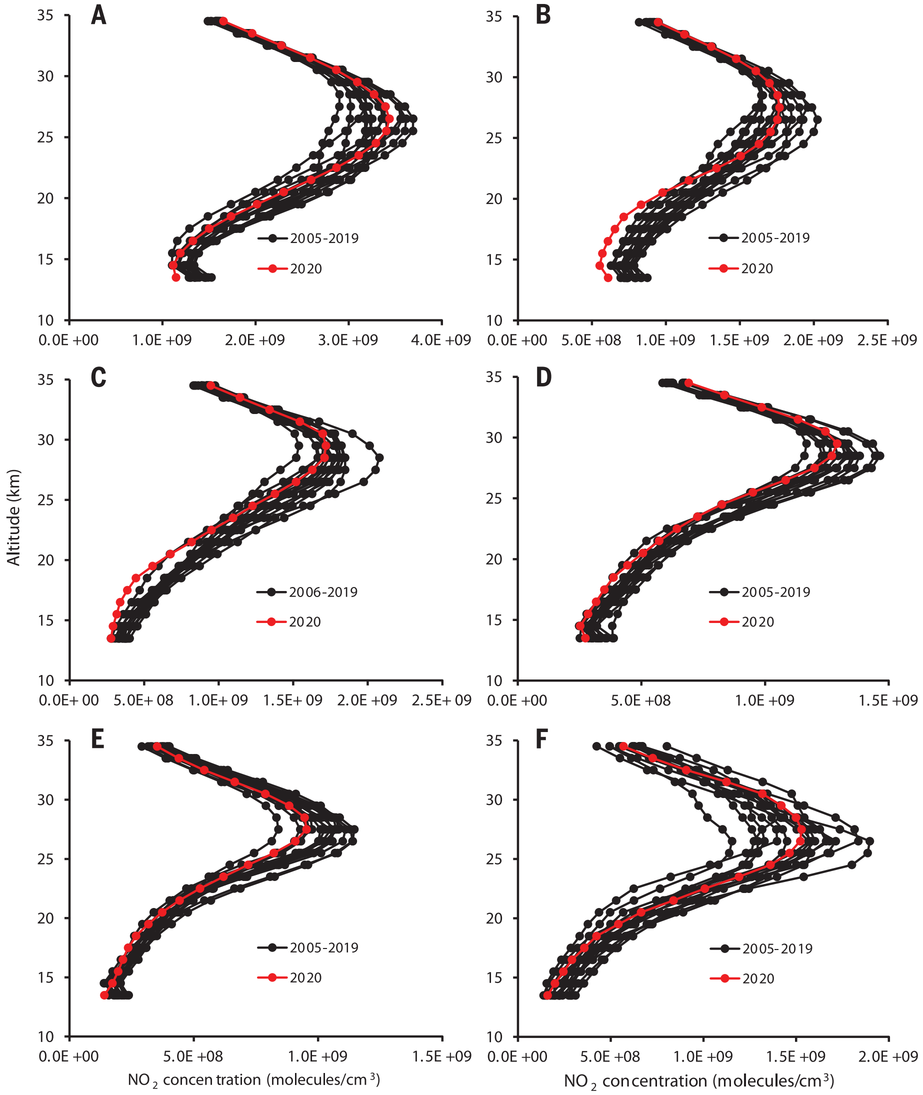


Fig. 4. NO_2 concentration profiles for ACE-FTS measurements in the latitude range 45° to 60°S. Average profiles for all years before 2020 are shown in black and the average profile for 2020 is in red. (**A**) January. (**B**) March. (**C**) April. (**D**) Late May and early June. (**E**) July. (**F**) August.

The strong enhancement of HOCl concentrations is particularly notable. HOCl is produced from ClO and $ClONO_2$. At mid-latitudes the reaction making HOCl (*20*, *21*) is

$$ClO + HO_2 \rightarrow HOCl + O_2 \qquad (1)$$

The injection of organic molecules into the stratosphere leads to the formation of formaldehyde (H_2CO) and other sources of HO_2 radicals (*18*); H_2CO near 18 km was observed to be enhanced well into 2021 (fig. S4). Photolysis of H_2CO or reaction with hydroxyl radical (OH) leads to the formyl radical (HCO), which reacts with O_2 to make excess hydroperoxyl radical (HO_2). HO_2 depletes O_3 (*19*) in the HO_x cycle ($OH + O_3 \rightarrow HO_2 + O_2$, $HO_2 + O_3 \rightarrow OH + 2O_2$) and forms HOCl in reaction 1. HOCl also depletes ozone when it is photolyzed to make OH and Cl, both of which destroy O_3 in the HO_x and ClO_x ($Cl + O_3 \rightarrow ClO + O_2$, $ClO + O \rightarrow Cl + O_2$) cycles (*19*).

HOCl production (Fig. 3) from $ClONO_2$ is also likely catalyzed by hydrated smoke particles,

$$ClONO_2 + H_2O \rightarrow HOCl + HNO_3 \qquad (2)$$

Heterogeneous chemistry of HOCl on smoke then can contribute to the observed reduction (Fig. 3) in HCl abundance through the reaction

$$HOCl + HCl \rightarrow Cl_2 + H_2O \qquad (3)$$

The heterogeneous reaction of $ClONO_2$+HCl also yields Cl_2 and is likely to contribute to the loss of HCl. Cl_2 is photolyzed to Cl atoms, and ozone destruction occurs through the ClO_x cycle. Reactions 2 and 3 are known to be catalyzed by polar stratospheric clouds (PSCs) during Antarctic and Arctic ozone depletion events (*18*). Similar chemistry can be expected on hydrated acidic smoke particles. The 2020 time series for $ClONO_2$, HCl, HOCl, and ClO concentration profiles are presented in figs. S5, S6, S7, and S8, respectively.

Nitrogen dioxide (NO_2) abundances (Fig. 4) were reduced in the 15- to 25-km range at mid-latitudes from January to July 2020 (particularly in March and April). Dinitrogen pentoxide (N_2O_5) forms at night from NO_2 and reacts with H_2O in aerosol particles (*22*),

$$N_2O_5 + H_2O \rightarrow 2HNO_3 \qquad (4)$$

This reaction is slow in the gas phase but may be catalyzed by hydrated smoke particles analogous to volcanic sulfate aerosol chemistry (*23*). Galib and Limmer (*24*) have recently shown that the uptake of N_2O_5 on aqueous aerosols is a surface process, supporting the idea that N_2O_5 reacts on the hydrated surface of smoke particles. The variability of the NO_2 profiles in other years is likely a result of stratospheric dynamics (Fig. 4).

$ClONO_2$ is produced by three-body recombination of ClO and NO_2 (*19*). The observed enhancement (Fig. 3) of $ClONO_2$ means that the increased production of the reactive ClO radical (modest enhancements are observed by ACE-FTS; Fig. 3) offsets the decline in NO_2. The perturbation of HCl by smoke is less extreme than that caused by polar stratospheric clouds (PSCs) during polar winter, when HCl abundances can approach zero (*19*). The observed stratospheric temperatures from ACE-FTS at mid-latitudes are never low enough to produce PSCs. Indeed, the average temperature profiles were elevated from January to March 2020 (fig. S9), consistent with stratospheric heating by smoke particles (*7*). Polar chlorine chemistry strongly reduces the abundances of both HCl and $ClONO_2$ by reactions on PSCs in contrast to the reduction in HCl and the increase in $ClONO_2$ associated with smoke particles.

ACE-FTS observations clearly show that pyroCb organics and smoke cause substantial perturbations in chemical species, demonstrating an unanticipated role in stratospheric chemistry. Although the reductions in NO_x will increase O_3 abundances by suppression of the NO_x cycle (*18*) ($NO + O_3 \rightarrow NO_2 + O_2$, $NO_2 + O \rightarrow NO + O_2$), this effect is offset by enhanced O_3 destruction in the ClO_x and HO_x cycles. As severe wildfires rise in number, they will play an increasingly important role in the global ozone budget.

REFERENCES AND NOTES

1. IPCC, "Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change," V. Masson-Delmotte *et al.*, Eds. (Cambridge Univ. Press, 2021); www.ipcc.ch/report/ar6/wg1/

2. N. Glatthor *et al.*, *Atmos. Chem. Phys.* **13**, 1637–1658 (2013).
3. S. M. Khaykin *et al.*, *Geophys. Res. Lett.* **45**, 1639–1646 (2018).
4. D. A. Peterson *et al.*, *NPJ Clim. Atmos. Sci.* **4**, 38 (2021).
5. M. Fromm *et al.*, *Bull. Am. Meteorol. Soc.* **91**, 1193–1210 (2010).
6. C. D. Boone, P. F. Bernath, M. D. Fromm, *Geophys. Res. Lett.* **47**, e2020GL088442 (2020).
7. P. Yu *et al.*, *Geophy. Res. Lett.* **48**, e2021GL092609 (2021).
8. G. Di Virgilio *et al.*, *Geophys. Res. Lett.* **46**, 8517–8526 (2019).
9. D. A. Peterson *et al.*, *NPJ Clim. Atmos. Sci.* **1**, 30 (2018).
10. S. Solomon *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2117325119 (2022).
11. P. F. Bernath, *J. Quant. Spectrosc. Radiat. Transf.* **186**, 3–16 (2017).
12. UN Environment Programme, "Montreal protocol on substances that deplete the ozone layer" (1987); https://ozone.unep.org/treaties/montreal-protocol.
13. World Meteorological Organization, "Scientific Assessment of Ozone Depletion: 2018, Global Ozone Research and Monitoring Project" (Report No. 58, pp. 588, 2018); https://csl.noaa.gov/assessments/ozone/2018/.
14. P. Bernath, A. M. Fernando, *J. Quant. Spectrosc. Radiat. Transf.* **217**, 126–129 (2018).
15. M. Amos *et al.*, *Atmos. Chem. Phys.* **20**, 9961–9977 (2020).
16. C. D. Boone, P. F. Bernath, D. Cok, J. Steffen, S. C. Jones, *J. Quant. Spectrosc. Radiat. Transf.* **247**, 106939 (2020).
17. P. F. Bernath, J. Steffen, J. Crouse, C. D. Boone, *J. Quant. Spectrosc. Radiat. Transf.* **253**, 107178 (2020).
18. B. J. Finlayson-Pitts, J. N. Pitts Jr., *Chemistry of the Upper and Lower Atmosphere: Theory, Experiments, and Applications* (Academic Press, 1980).
19. S. Solomon, *Rev. Geophys.* **37**, 275–316 (1999).
20. T. von Clarmann *et al.*, *Atmos. Chem. Phys.* **12**, 1965–1977 (2012).
21. P. F. Bernath, R. Dodandodage, C. D. Boone, J. Crouse, *J. Quant. Spectrosc. Radiat. Transf.* **264**, 107559 (2021).
22. T. H. Bertram, J. A. Thornton, *Atmos. Chem. Phys.* **9**, 8351–8363 (2009).
23. M. Prather, *J. Geophys. Res.* **97**, 10,187–10,191 (1992).
24. M. Galib, D. T. Limmer, *Science* **371**, 921–925 (2021).

ACKNOWLEDGMENTS

P.B. acknowledges R.F. Bernath for productive discussion. **Funding:** Canadian Space Agency contract 9F045-200575/001/SA (P.B., C.B., and J.C.) **Author contributions:** Conceptualization: P.B. and C.B. Methodology: C.B. and P.B. Visualization: C.B. and J.C. Funding acquisition: P.B. Supervision: P.B. Writing – original draft: P.B. Writing – review and editing: P.B., C.B., and J.C. **Competing interests:** Authors declare that they have no competing interests. **Data and materials availability:** Data for Fig. 1 are available in the supplementary materials. ACE data, including the HOCl research product, are freely available after sign-up at https://databace.scisat.ca/l2signup.php.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm5611
Materials and Methods
Figs. S1 to S9
Reference (*25*)
Data for Fig. 1

25 September 2021; accepted 11 February 2022
10.1126/science.abm5611

GRAPHENE

Spectroscopy signatures of electron correlations in a trilayer graphene/hBN moiré superlattice

Jixiang Yang[1]†, Guorui Chen[2,3,4]†, Tianyi Han[1]†, Qihang Zhang[1], Ya-Hui Zhang[5], Lili Jiang[4], Bosai Lyu[2], Hongyuan Li[3,4], Kenji Watanabe[6], Takashi Taniguchi[6], Zhiwen Shi[2], Todadri Senthil[1], Yuanbo Zhang[7,8], Feng Wang[3,4,9], Long Ju[1]*

ABC-stacked trilayer graphene/hexagonal boron nitride moiré superlattice (TLG/hBN) has emerged as a playground for correlated electron physics. We report spectroscopy measurements of dual-gated TLG/hBN using Fourier transform infrared photocurrent spectroscopy. We observed a strong optical transition between moiré minibands that narrows continuously as a bandgap is opened by gating, indicating a reduction of the single-particle bandwidth. At half-filling of the valence flat band, a broad absorption peak emerges at ~18 milli–electron volts, indicating direct optical excitation across an emerging Mott gap. Similar photocurrent spectra are observed in two other correlated insulating states at quarter- and half-filling of the first conduction band. Our findings provide key parameters of the Hubbard model for the understanding of electron correlation in TLG/hBN.

Moiré superlattices of two-dimensional (2D) materials (*1*–*4*) offer a versatile platform to engineer and study correlated electron physics. With additional experimental knobs such as the composition, the twisting angle, and gate electric fields, the electronic band structure and charge density can be (in situ) independently controlled in these synthetic quantum materials. As a result, superconductivity (*5*–*10*), correlated insulating states (*11*–*18*), orbital magnetism (*19*), and correlated Chern insulators (*12*, *20*) have been observed in 2D moiré superlattices. Many intriguing questions arise from these observations, such as the nature of correlated insulating states and the mechanism of superconductivity, which call for systematic spectroscopy studies. Although several spectroscopy studies have been carried out (*21*–*26*) on the particular system of magic-angle twisted bilayer graphene (MATBLG), experiments on 2D moiré superlattices have been dominated by electron transport measurements. Applying additional top layers for encapsulation (to mimic the dielectric screening of the environment, as in transport experiments) or dual gating (*13*–*17*, *27*) (relevant to the majority of 2D moiré superlattices) makes it impossible to use many of the spectroscopic techniques that are commonly used in surface science, such as scanning tunneling spectroscopy and photoemission spectroscopy. To gain a fundamental understanding of the correlated physics in 2D moiré superlattices, developing spectroscopy techniques that can access buried heterostructures is of critical importance at this exciting frontier of research.

Optical transmission and reflection spectroscopy has been widely used to study electronic excitations in conventional strongly correlated materials (*28*). However, in moiré superlattices, the relevant energy scales of the electronic band and Coulomb gap are substantially smaller (<50 meV) than their counterparts in conventional strongly correlated materials. Consequently, far-infrared spectroscopy with photon wavelengths >25 μm is required to probe correlated electron excitations in moiré superlattices. The corresponding beam spot size (~1 mm when using Globar, the only broadband infrared source) is much larger than that of typical 2D material devices (*29*), making infrared reflection and transmission spectroscopy of moiré heterostructures extremely challenging. At the same time, infrared absorption in the gate layers easily dominates that from the layer of interest, making it harder to extract a clean signal. So far, no infrared spectroscopy has been applied to top-gated 2D moiré superlattices.

Here, we overcome these challenges by adopting Fourier transform infrared (FTIR)–photocurrent spectroscopy for the optical absorption measurement of an ABC-stacked trilayer graphene/hexagonal boron nitride moiré superlattice (TLG/hBN). The overall methodology is similar to that described in (*29*). However, a notable improvement of signal-to-noise ratio has been achieved in this study, which enables spectroscopy of correlated states in moiré superlattices (*30*); we applied this technique to dual-gated TLG/hBN (as illustrated in Fig. 1A).

Figure 1B shows the calculated band structure of TLG/hBN in the mini-Brillouin zone using a tight-binding model, where a moiré potential exists at the top layer of TLG (*30*). Dashed curves represent the four lowest moiré

---

[1]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA, USA. [2]Key Laboratory of Artificial Structures and Quantum Control (Ministry of Education), Shenyang National Laboratory for Materials Science, School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, China. [3]Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [4]Department of Physics, University of California at Berkeley, Berkeley, CA, USA. [5]Department of Physics, Harvard University, Cambridge, MA, USA. [6]National Institute for Materials Science, Tsukuba, Japan. [7]State Key Laboratory of Surface Physics and Department of Physics, Fudan University, Shanghai, China. [8]Institute for Nanoelectronic Devices and Quantum Computing, Fudan University, Shanghai, China. [9]Kavli Energy NanoSciences Institute at the University of California, Berkeley, CA, USA.
***Corresponding author. Email: longju@mit.edu**
†These authors contributed equally to this work.

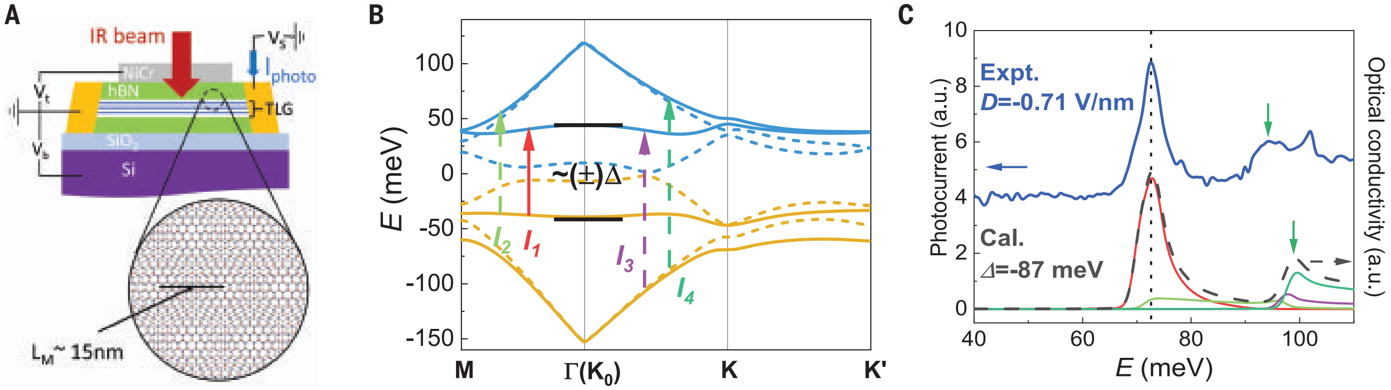


Fig. 1. Device structure and interband optical transitions in an ABC-stacked TLG/hBN moiré superlattice. (**A**) Illustration of a dual-gated TLG device with a moiré wavelength of ~15 nm. IR, infrared; V_b, bottom-gate voltage; V_t, top-gate voltage; V_s, source-drain voltage. (**B**) Calculated band structure of TLG in the mini-Brillouin zone under a moiré potential (*30*). Dashed curves represent minibands at zero displacement field, whereas solid curves indicate minibands at a displacement field induced by gates so that $\Delta = -87$ meV. Δ represents the gate-induced potential difference between the top and bottom layer of TLG. Arrows labeled as I_1 to I_4 represent interband optical transitions when the Fermi level is inside the gap. (**C**) The photocurrent spectrum of TLG/hBN at a displacement field $D = -0.71$ V/nm and calculated optical conductivity spectrum at $\Delta = -87$ meV. Calculated contributions from transitions I_1 to I_4 are plotted as solid curves, and the sum is plotted as a dashed curve. The prominent experimental peak at ~72 meV (indicated by the vertical dashed line) corresponds to transition I_1. The peak in the photocurrent spectrum indicated by a green arrow corresponds to the peak in optical conductivity that is dominated by I_3 and I_4. All photocurrent spectroscopy measurements were performed at a sample temperature of 2 K. a.u., arbitrary units; Cal., calculated; Expt., experimental.

minibands at zero displacement field D, where the bandgap at charge neutrality is zero. As we apply a positive (negative) D by controlling bottom- and top-gate voltages, a positive (negative) potential energy difference between the bottom and top graphene layers is induced ($\Delta = E_t - E_b$); the resulting minibands are shown as solid curves. Here, we define the positive direction of D to be pointing from the bottom gate to the top gate. This potential energy asymmetry corresponds to a bandgap that is slightly smaller than $|\Delta|$. The highest valence band is well separated from the other bands, forming an obvious flat band that hosts correlated insulating states when doped (*27*, *31*, *32*). The lowest conduction band also features a small bandwidth, but it partially overlaps with the second-lowest conduction band. Interband optical transitions can happen between these moiré minibands, as indicated by arrows I_1 to I_4. I_1 and I_4 have the largest oscillator strengths because they are allowed even in the absence of the moiré potential. By contrast, I_2 and I_3 are allowed only by the moiré potential effect, and they contribute less to optical conductivity. Figure 1C shows the experimental photocurrent spectrum at $D = -0.71$ V/nm, which features a sharp and strong peak at ~72 meV. The photocurrent signal is zero below this peak. There is a second broader peak residing at ~20 meV above the main peak. The sharp peak at ~102 meV is caused by interlayer electron-phonon coupling between hBN and graphene and requires further study.

Using the calculated optical conductivity spectrum in Fig. 1C as a reference, the experimental photocurrent spectrum can be interpreted. The main peak at ~72 meV is dominated by the interband transition I_1 in Fig. 1B, and the second peak at ~95 meV may result from I_3 and I_4 transitions. The separation between these two peaks corresponds to the bandgap between the two valence bands, supporting the important presumption that the highest valence band is isolated when considering correlation effects. The peak width of the main peak is determined by the joint density of states (DOS), which is closely related to the peak width of the van Hove singularities in the DOS (*30*) of the highest valence band and the lowest conduction band. When both bands are relatively flat as in bandgap-opened TLG, the observed peak width serves as a good indicator of the single-particle bandwidth W.

In contrast to MATBLG, where W is determined mostly by the twisting angle, theory predicts that the moiré bandwidth in TLG/hBN can be tuned in situ by D (*32*). Our measurement allows a direct visualization of this electrical tuning of the moiré flatband bandwidth. Figure 2A shows photocurrent spectra of TLG/hBN at several values of D, where the charge density is fixed at zero. At $D = -0.38$ V/nm, the spectrum features a broad peak at ~35 meV with a full width at half-maximum (FWHM) of ~20 meV. This wide peak width is caused by the relative dispersive bands at small displacement fields. As D changes from -0.38 to -0.55 to -0.71 V/nm, this peak blue-shifts and quickly sharpens monotonically. At positive D and Δ, a similar broad peak is observed at $D = 0.33$ V/nm. This peak sharpens at $D = 0.49$ V/nm and eventually broadens again at $D = 0.66$ V/nm, showing a nonmonotonic change of peak width. At the same time, a second peak corresponding to I_3 and I_4 emerges gradually as $|D|$ increases in both directions. Such trends are better seen in the color plot of the photocurrent spectra in Fig. 2B when D is continuously tuned. Dashed lines correspond to the spectra in Fig. 2A with the same color. We calculated optical conductivity spectra as a function of D, as shown in Fig. 2C. By aligning the lowest-energy peak position in both the experimental photocurrent spectrum and the calculated optical conductivity spectrum, we established a conversion from D to Δ (*30*). This allowed us to outline two regions by white boxes in Fig. 2C that correspond to the experimental data in Fig. 2B. Major features and trends in Fig. 2B are well reproduced by calculation in Fig. 2C. The blue shift of the I_1-dominated optical transition peak indicates the opening of the bandgap. At the same time, the bandwidth of the relevant moiré bands continuously evolves.

Figure 2D summarizes the FWHM of the I_1-dominated peak as a function of Δ. Both experimental and calculated FWHM decrease monotonically to ~5 meV as Δ moves in the negative direction. With $\Delta > 0$, a minimum FWHM of ~7 meV is observed experimentally at $\Delta \sim 60$ meV. The calculated peak width of the DOS of the highest valence band (*30*) is plotted as blue diamonds. It shows a trend similar to that of the optical transition peak FWHM, reaching a minimum of 5 meV. The evolution of the optical transition peak width largely reflects that of the bandwidth W of the highest valence band, which is plotted as green triangles. As the bandgap is opened, the bandwidth W is suppressed and eventually approaches ~12 meV. This value is smaller than

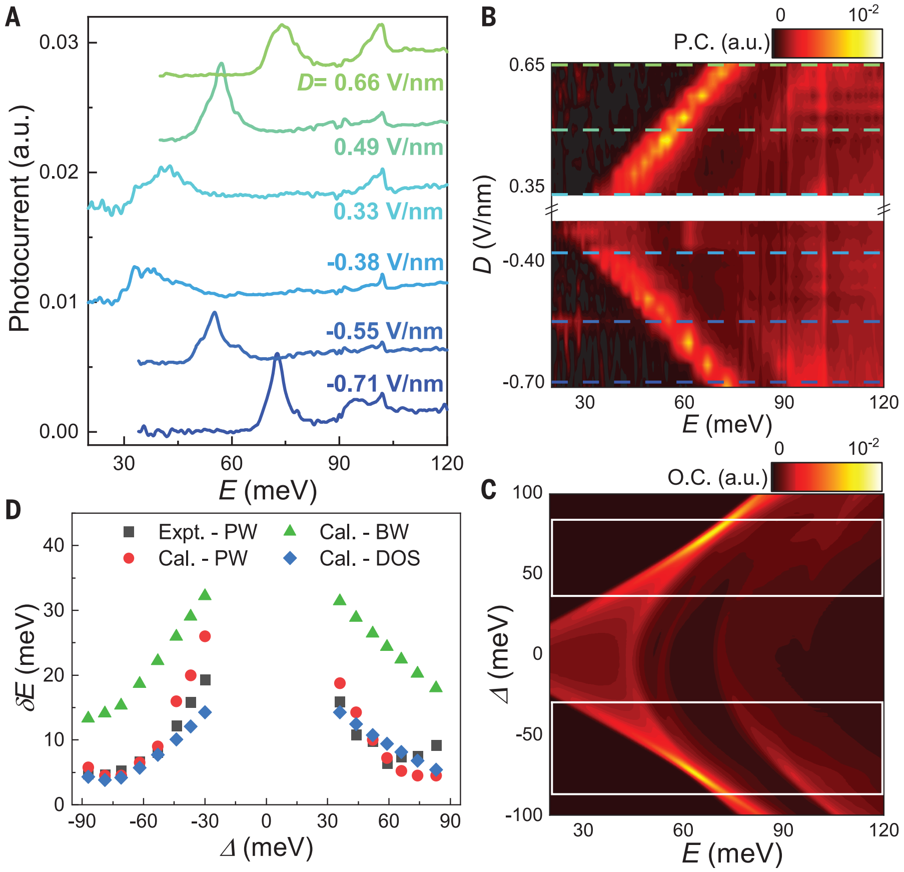


Fig. 2. Displacement field–dependent interband optical transitions at zero doping. (**A**) Photocurrent spectra at several representative displacement fields and zero doping at 2 K. The spectra are shifted vertically for clarity. At *D* = −0.38 V/nm, the low-energy peak is very broad in energy, indicating a large dispersion of both conduction and valence bands. As *D* becomes more negative and the bandgap increases, this peak blue-shifts and narrows, indicating that the lowest conduction and valence bands are being squeezed by the bandgap. At positive *D*, the peak position also blue-shifts monotonically, whereas the peak width first narrows and then broadens. (**B**) The 2D color plot of the photocurrent (P.C.) spectra as a function of *D*. Dashed lines correspond to the spectra in (A) with the same colors. (**C**) The 2D color plot of the calculated optical conductivity (O.C.) spectra as a function of Δ. The two white boxes outline the corresponding range of data in (B). A good agreement with (B) is observed for both I_1- and I_4-dominated spectrum ranges. (**D**) FWHM of the I_1-dominated peak as a function of Δ extracted from (B) and (C) are shown as black squares and red dots, respectively. Blue diamonds represent the calculated DOS van Hove singularities peak width of the highest valence band. Green triangles represent the full bandwidth of the highest valence band. As the bandgap is opened, both single-particle DOS peak and optical transition peak widths become considerably narrower, driving the system into a correlated-electron regime when doped. BW, bandwidth; PW, peak width.

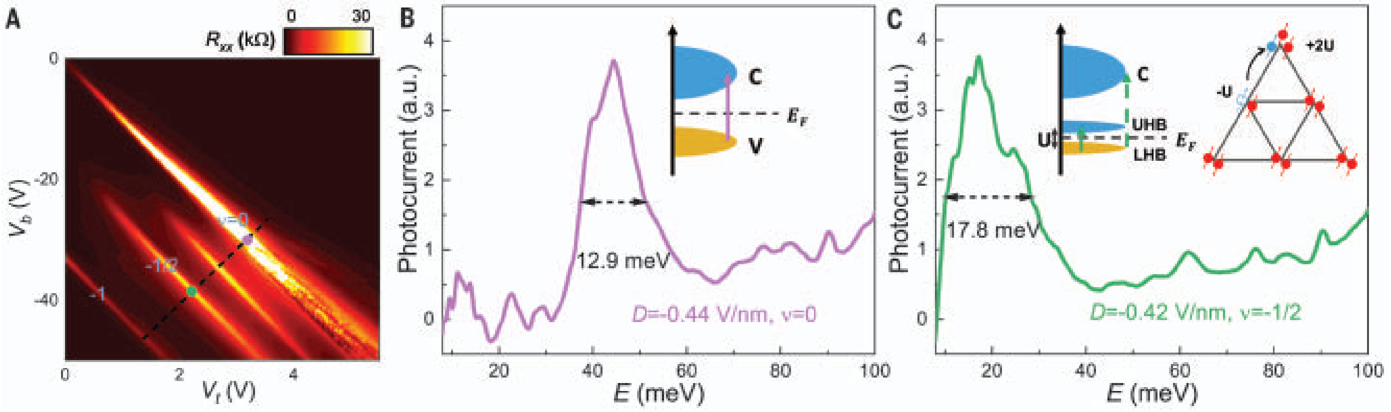


Fig. 3. Optical transitions in the correlated insulating state. (**A**) Device resistance (*R*) as a function of the top- and bottom-gate voltages. The dashed line indicates the constant *D* direction. Resistance peaks corresponding to full-filling, half-filling, and zero-filling of the highest valence band are labeled as ν = −1, −½, and 0, respectively. (**B**) The photocurrent spectrum taken at the purple dot position in (A). The spectrum is dominated by a sharp peak at ~45 meV, which corresponds to the I_1 transition, as illustrated by the inset. C, conduction band; E_F, Fermi energy; V, valence band. (**C**) Photocurrent spectrum taken at the green dot position in (A). In contrast to the spectrum in (B), a broad peak at ~18 meV emerges, corresponding to an optical transition across the Mott gap, as illustrated by the solid arrow in the inset on the left. The final state of such optical excitation contains a hole at one site and an extra electron at another site in the triangular moiré superlattice, as illustrated in the inset on the right. The I_1 peak (indicated by the dashed arrow in the inset on the left) is merged into the background. LHB, lower Hubbard band; UHB, upper Hubbard band. The spectra in (B) and (C) are taken at 2 K.

the estimated on-site Coulomb repulsion energy $U = \frac{e^2}{4\pi\varepsilon_0\varepsilon l_M} \approx 25$ meV (where $l_M = 15$ nm is the moiré wavelength, and ε = 4 is the dielectric constant of hBN), making correlation effects possible when partial doping is induced by gates. Our data also indicated a slightly broader peak width at Δ > 0 than at Δ < 0 for the whole range (*30*). This observed asymmetry of peak width implies a bigger *W* for the Δ > 0 side, which agrees with the fact that correlation effects are stronger at negative *D* than at positive *D* in transport experiments (*8*) as well as with the theoretical calculation of bandwidth *W*, as shown in Fig. 2D.

Next, we examine optical transitions in the correlated insulating state when the flat valence band is doped. Figure 3A shows the device resistance as a function of top- and bottom-gate voltages. At half-filling of the flat valence band (ν = −½), a correlated insulating state is formed, as indicated by a resistance peak. Figure 3B shows the photocurrent spectrum at zero-filling of this hole band at D = −0.44 V/nm, where a photocurrent peak centered at 45 meV corresponds to I_1, as illustrated by the inset. This peak has a FWHM of ~13 meV; the calculations in Fig. 2D predict a DOS peak width of 10 meV at this D. At ν = −½ with a similar displacement field D = −0.42 V/nm, the photocurrent spectrum is dramatically changed, as shown in Fig. 3C. We observed a new strong peak centered at ~18 meV. This energy is clearly below the bandgap energy in Fig. 3B, yet it is bigger than the DOS peak width of the flat valence band. The FWHM of this new peak is ~18 meV, which is substantially broader than the peak width in Fig. 3B. The I_1-dominated peak is merged into the broad background.

These observations at ν = −½ all point to a picture that is much more complicated than that in a doped band insulator, where electron correlation effects can be neglected. For a half-doped single-band Hubbard model, theoretical calculations (*33*–*35*) predicted a broad peak in the optical conductivity spectrum centering around on-site Coulomb repulsion energy U. Such features have been observed by optical spectroscopy experiments in conventional strongly correlated materials (*33*, *36*, *37*). We believe that the strong peak at 18 meV in Fig. 3C indicates the formation of upper and lower Hubbard bands with an optical excitation across the Mott gap, as illustrated by the inset of Fig. 3C. In the final state of this optical excitation, a hole is left at one site while an extra electron is added to another site on the triangular moiré superlattice of TLG/hBN. For the moiré superlattice in our device, $\frac{e^2}{4\pi\varepsilon_0\varepsilon l_M} \approx 25$ meV is expected to be a good estimate of the onsite Coulomb repulsion energy U (*38*). This energy is close to the peak position of 18 meV in Fig. 3C. The second piece of evidence of correlation effects is the large broadening of the low-energy peak. The FWMH of 18 meV is several times bigger than expected from the simple uncorrelated band picture: Without correlations, one expects a peak width of ~6.5 meV, as inferred from the experimental spectrum in Fig. 3B, and ~5 meV from calculations. Thirdly, the interband transition from the lower Hubbard band to the lowest conduction miniband (illustrated by the dashed arrow in the inset of Fig. 3C) is difficult to distinguish from the broad continuous background. It is likely that the DOS distribution of the lowest conduction band, which is remote to the flat valence band, also gets substantially broadened and prevents an easy identification of the transition from the experimental spectrum. The broadening of both the flat band and remote band is similar to what happens in MATBLG (*21*, *22*, *24*), indicating that strong correlation effects play a key role in our system.

We now briefly discuss the implication of this low-energy optical transition on the detailed nature of the Mott insulator state at ν = −½ of the valence flat band. We can rule out spin- and valley-polarized ferromagnetic ground states, because optical transitions are forbidden in such states owing to the conservation of spin and valley pseudospin. This is also supported by transport measurements where magnetism is absent from the topologically trivial side of TLG/hBN (*12*). Other candidate ground states such as antiferromagnetic and intervalley-coherent Mott insulators on a triangular lattice (*36*, *38*) allow optical transitions across the charge gap. They both agree with our experimental results, but a more precise identification requires further experimental and theoretical studies. In particular, continuous tuning of Δ and intersite coupling parameters in TLG/hBN facilitate experimental exploration of multiple ground states.

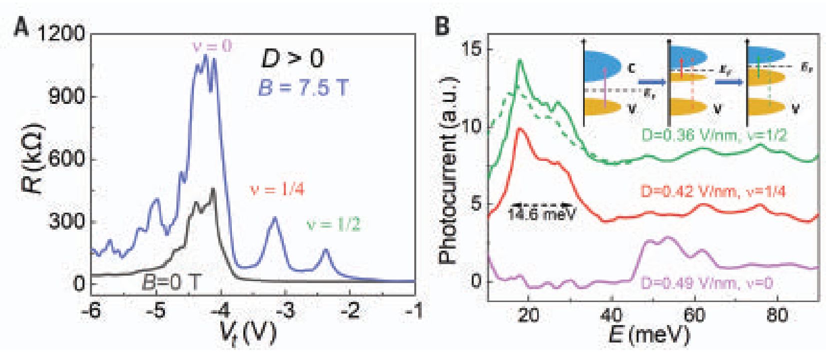


Fig. 4. Optical transitions in magnetic field–induced correlated insulating states. (**A**) Top gate–dependent device resistance at a fixed bottom-gate voltage V_b = +38 V. At zero magnetic field, a single resistance peak appears at filling factor ν = 0. At B = 7.5 T, additional peaks appear at ν = ¼ and ½, corresponding to quarter- and half-fillings of the lowest conduction band. (**B**) Photocurrent spectra corresponding to the three resistance peaks at 7.5 T in (A). The ν = ¼ and ½ spectra are shifted by 3 and 6 a.u. respectively for clarity. At charge density n = 0, a peak centered at 50 meV indicates the optical transition across the main gap. At both ν = ¼ and ½, a peak centered at around ~22 meV appears, whereas the interband transition is merged into background. This low-energy peak mimics the peak in Fig. 3C (shown as a dashed curve for comparison) in both peak position and peak width, implying similar band-splitting physics as that in a Mott insulator. All the transport data in (A) and optical spectra in (B) are measured at 2 K.

Besides the correlated insulating state at ν = −½ of the flat valence band, we also explored magnetic field–induced insulating states when the Fermi level is shifted into the lowest conduction band at D > 0. Figure 4A shows the transport signature of such states when D is positive. At zero magnetic field, the device resistance features a single peak corresponding to filling factor ν = 0, whereas the electron-doped side is featureless. However, at magnetic field B = 7.5 T, resistance peaks appear at ν = ¼ and ½ of the lowest conduction band. These features mimic the transport signatures of correlated insulating states at −¼ and −½ filling of the flat valence band when D is negative. Figure 4B shows the photocurrent spectra corresponding to these two new insulating states. At ν = 0, a peak located at ~50 meV indicates interband transition across the displacement field–induced bandgap. By contrast, at quarter- and half-filling, a peak located at ~22 meV emerges and dominates both spectra; no obvious peak can be identified near 50 meV at quarter- and half-filling.

Because ν = ¼ and ½ correspond to one electron and two electrons per site on the moiré superlattice, we believe that these two states are also driven by electron correlation, as in the Mott insulator state of the flat valence band at ν = −½. The exact role of the magnetic field is to be clarified by further measurements. Phenomenologically, the peak positions and widths in photocurrent spectra at these two magnetic field–induced insulating states are similar to those of the peak in Fig. 3C. We believe that a correlation-driven band splitting similar to that in Fig. 3C is present in this scenario, as illustrated by the inset of Fig. 4B. In both cases, the optical excitation energy is largely determined by the onsite Coulomb repulsion energy U, which is set by the same moiré wavelength and independent of the displacement field and the doping type.

Our measurements provide spectroscopic evidence of electron correlation effects in a TLG/hBN moiré superlattice. We have experimentally determined energy scales for relevant parameters of the Hubbard model, which form a basis for accurate theoretical modeling and understanding of both observed (*8*, *12*, *27*) and predicted correlated ground states (*36*–*38*) in this moiré superlattice. These observations open up opportunities to explore the doping and temperature dependence of the optical spectrum, sum rules for optical conductivity (*28*, *33*), and bound excitons of holon and doublon (*28*, *39*), which all call for further systematic study and theoretical calculations. The FTIR photocurrent spectroscopy technique used here can be readily generalized to other encapsulated and (dual-)gated 2D moiré superlattice devices for a better understanding of correlated electron physics in this designer material platform.

REFERENCES AND NOTES

1. M. Yankowitz *et al.*, *Nat. Phys.* **8**, 382–386 (2012).
2. C. R. Dean *et al.*, *Nature* **497**, 598–602 (2013).
3. L. A. Ponomarenko *et al.*, *Nature* **497**, 594–597 (2013).
4. B. Hunt *et al.*, *Science* **340**, 1427–1430 (2013).
5. Y. Cao *et al.*, *Nature* **556**, 43–50 (2018).
6. M. Yankowitz *et al.*, *Science* **363**, 1059–1064 (2019).
7. X. Lu *et al.*, *Nature* **574**, 653–657 (2019).
8. G. Chen *et al.*, *Nature* **572**, 215–219 (2019).
9. J. M. Park, Y. Cao, K. Watanabe, T. Taniguchi, P. Jarillo-Herrero, *Nature* **590**, 249–255 (2021).
10. Z. Hao *et al.*, *Science* **371**, 1133–1138 (2021).
11. Y. Cao *et al.*, *Nature* **556**, 80–84 (2018).
12. G. Chen *et al.*, *Nature* **579**, 56–61 (2020).
13. X. Liu *et al.*, *Nature* **583**, 221–225 (2020).
14. Y. Cao *et al.*, *Nature* **583**, 215–220 (2020).
15. S. Chen *et al.*, *Nat. Phys.* **17**, 374–380 (2021).
16. E. C. Regan *et al.*, *Nature* **579**, 359–363 (2020).
17. Y. Tang *et al.*, *Nature* **579**, 353–358 (2020).
18. L. Wang *et al.*, *Nat. Mater.* **19**, 861–866 (2020).
19. A. L. Sharpe *et al.*, *Science* **365**, 605–608 (2019).
20. M. Serlin *et al.*, *Science* **367**, 900–903 (2020).
21. Y. Xie *et al.*, *Nature* **572**, 101–105 (2019).
22. A. Kerelsky *et al.*, *Nature* **572**, 95–100 (2019).
23. Y. Jiang *et al.*, *Nature* **573**, 91–95 (2019).
24. Y. Choi *et al.*, *Nat. Phys.* **15**, 1174–1180 (2019).
25. M. I. B. Utama *et al.*, *Nat. Phys.* **17**, 184–188 (2021).
26. S. Lisi *et al.*, *Nat. Phys.* **17**, 189–193 (2021).
27. G. Chen *et al.*, *Nat. Phys.* **15**, 237–241 (2019).
28. D. N. Basov, R. D. Averitt, D. van der Marel, M. Dressel, K. Haule, *Rev. Mod. Phys.* **83**, 471–541 (2011).
29. L. Ju *et al.*, *Science* **358**, 907–910 (2017).
30. See additional text and data in supplementary materials.
31. B. L. Chittari, G. Chen, Y. Zhang, F. Wang, J. Jung, *Phys. Rev. Lett.* **122**, 016401 (2019).
32. Y. H. Zhang, D. Mao, Y. Cao, P. Jarillo-Herrero, T. Senthil, *Phys. Rev. B* **99**, 075127 (2019).
33. M. Imada, A. Fujimori, Y. Tokura, *Rev. Mod. Phys.* **70**, 1039–1263 (1998).
34. E. Dagotto, *Rev. Mod. Phys.* **66**, 763–840 (1994).
35. W. Stephan, P. Horsch, *Phys. Rev. B* **42**, 8736–8739 (1990).
36. H. C. Po, L. Zou, A. Vishwanath, T. Senthil, *Phys. Rev. X* **8**, 031089 (2018).
37. C. Xu, L. Balents, *Phys. Rev. Lett.* **121**, 087001 (2018).
38. Y. H. Zhang, T. Senthil, *Phys. Rev. B* **99**, 205150 (2019).
39. J. van den Brink, M. B. J. Meinders, J. Lorenzana, R. Eder, G. A. Sawatzky, *Phys. Rev. Lett.* **75**, 4658–4661 (1995).
40. J. Yang, Replication data for: Spectroscopy signatures of electron correlations in a trilayer graphene/hBN moiré superlattice (abg3036), Version 1. Harvard Dataverse (2022); https://doi.org/10.7910/DVN/LMBGXF.

ACKNOWLEDGMENTS

We acknowledge discussions with P. Lee, L. Levitov, R. Ashoori, L. Fu, P. Jarillo-Herrero, and K. F. Mak. **Funding:** The photocurrent spectroscopy measurement of this work was supported by the STC Center for Integrated Quantum Materials, NSF grant no. DMR-1231319. The device fabrication was partially supported by the Skolkovo Institute of Science and Technology as part of the MIT Skoltech Program. G.C. and F.W. were supported as part of the Center for Novel Pathways to Quantum Coherence in Materials, an Energy Frontier Research Center funded by the US Department of Energy, Office of Science, Basic Energy Sciences. G.C. acknowledges financial support from the National Key Research Program of China (grant nos. 2020YFA0309000 and 2021YFA1400100), NSF of China (grant no.12174248), and Shanghai Jiao Tong University (no. 21X010200846). K.W. and T.T. acknowledge support from the Elemental Strategy Initiative conducted by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan, grant no. JPMXP0112101001; Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (KAKENHI) grant no. JP20H00354; and the CREST (JPMJCR15F3), Japan Science and Technology Agency (JST). B.L. and Z.S. acknowledge support from the National Key Research and Development Program of China (2016YFA0302001) and the National Natural Science Foundation of China (11774224 and 12074244). T.S. is supported by US Department of Energy grant DE-SC0008739 and in part by a Simons Investigator award from the Simons Foundation. **Author contributions:** L.Ju conceived and supervised the experiment. Q.Z., J.Y., and T. H. performed photocurrent spectroscopy measurements. G.C. and T.H. fabricated the sample with help from L.Ji., B.L., H.L., and Z.S. K.W. and T.T. grew hBN single crystals. J.Y., Y.-H.Z., and T.S. calculated the band structure and optical conductivity. J.Y., G.C., Q.Z., Y.Z., F.W., and L.Ju analyzed the data. J.Y. and L.Ju wrote the paper, with input from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All the data in the main text and supplementary materials, as well as the code to calculate the band structures and optical conductivity, can be obtained at Harvard Dataverse (*40*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abg3036
Methods
Supplementary Text
Figs. S1 to S11
References (*41*–*43*)

23 December 2020; accepted 16 February 2022
10.1126/science.abg3036

ULTRACOLD CHEMISTRY

Reactions between layer-resolved molecules mediated by dipolar spin exchange

William G. Tobias*, Kyle Matsuda, Jun-Ru Li, Calder Miller, Annette N. Carroll, Thomas Bilitewski, Ana Maria Rey, Jun Ye*

Microscopic control over polar molecules with tunable interactions enables the realization of distinct quantum phenomena. Using an electric field gradient, we demonstrated layer-resolved state preparation and imaging of ultracold potassium-rubidium molecules confined to two-dimensional planes in an optical lattice. The rotational coherence was maximized by rotating the electric field relative to the light polarization for state-insensitive trapping. Spatially separated molecules in adjacent layers interact through dipolar spin exchange of rotational angular momentum; by adjusting these interactions, we regulated the local chemical reaction rate. The resonance width of the exchange process vastly exceeded the dipolar interaction energy, an effect attributed to thermal energy. This work realized precise control of interacting molecules, enabling electric field microscopy on subwavelength scales and allowing access to unexplored physics in two-dimensional systems.

Ultracold polar molecules, which have complex internal structures and dipole moments tunable with external electric fields, represent a model system for studying many-body physics (*1*–*4*). In reduced dimensionality, the sign and magnitude of interactions between molecules depend on the orientation of the dipole moments with respect to the external confinement. Within two-dimensional (2D) layers, for example, the averaged interactions between molecules can be varied continuously from attractive to repulsive by rotating the dipoles into and out of the plane. Molecules in an isolated 2D layer are predicted to exhibit diverse quantum phenomena determined by the dipole angle and other parameters, including electric field and rotational state. These include complex ground state phases, such as superfluids and topological insulators (*5*–*12*), collective excitations in the hydrodynamic regime (*13*), and interaction-enhanced rotational coherence and dynamical generation of spin squeezing (*14*). Molecules prepared in multiple 2D layers may pair and form states with long-range order (*15*–*18*). Addressing individual lattice layers would allow initialization of varied configurations to realize these models—single layers, where molecules are isolated against out-of-plane interactions, and minimal systems with interlayer interactions, such as bilayers and trilayers (two and three adjacent layers, respectively).

Recent experimental progress with molecules in two dimensions has included reaching quantum degeneracy using direct evaporation (*19*) or pairing in a degenerate atomic gas (*20*), performing optical microscopy of individual lattice sites (*21*), and lengthening the coherence time of rotational superpositions (*22*, *23*). When translational motion is allowed within layers, as is the case for confinement in a 1D optical lattice, molecules approaching at short range undergo lossy chemical reactions (*24*–*28*). These losses can be mitigated by orienting the dipole moments perpendicular to the layer (*19*, *29*–*32*) or by engineering rotational state couplings

---

JILA, National Institute of Standards and Technology, and Department of Physics, University of Colorado, Boulder, CO 80309, USA.
*Corresponding author. Email: william.tobias@colorado.edu (W.G.T.); ye@jila.colorado.edu (J.Y.)

(*33–37*) to generate a repulsive collisional barrier. A major missing capability is the ability to prepare molecules in different internal states and control multiple layers individually, which permits local tuning of dipolar interactions in reduced dimensionality.

In this work, we demonstrated layer-resolved imaging and rotational state preparation of ultracold potassium-rubidium (KRb) molecules by applying an electric field gradient to shift rotational transition frequencies between lattice layers. This method was inspired by previous works with atoms (*38–40*). Using the capability to address single layers, we characterized the dynamics of highly controlled few-layer systems interacting by exchanging rotational angular momenta (referred to as spin exchange), a process mediated by dipolar interactions (*41*). Within individual layers, molecules experience both long-range dipolar interactions and short-range chemical reactions, and separated layers only interact through dipolar spin exchange. In the case of multiple layers containing different rotational states, spin exchange led to mixing of rotational state populations, which strongly enhanced the chemical reaction rate (*42*). Local control of molecule layer occupation and internal state allowed us to probe these dynamics and would potentially enable the exploration of quantum phases in two dimensions.

Starting with degenerate ^{40}K and ^{87}Rb atoms in an optical dipole trap (ODT), we loaded a 1D optical lattice to form a stack of 2D layers with an interlayer spacing a = 540 nm (*19*). Using magnetoassociation and stimulated Raman adiabatic passage (STIRAP) at a bias electric field $|\mathbf{E}| = 1\ \mathrm{kV\ cm^{-1}}$, we associated the atoms into KRb (*43*, *44*). KRb was formed in the rotational ground state $|0\rangle$, where $|N\rangle$ denotes the state with electric field–dressed rotational quantum number N and zero angular momentum projection ($m_N = 0$) onto the quantization axis specified by $\mathbf{E}$ (*45*). The harmonic trapping frequencies for $|0\rangle$ in the combined trap were (ω_x, ω_y, ω_z) = $2\pi \times (42, 17 \times 10^3, 48)$ Hz. The molecules were pinned along **y**, parallel to gravity, but they were free to move radially (**x**-**z** plane; Fig. 1A). For the typical temperature T = 350 nK, $k_B T = 0.4\ \hbar\omega_y$, where k_B is the Boltzmann constant and $\hbar = h/2\pi$ is the reduced Planck constant, so the molecules predominantly occupied the lowest lattice band. Because of the thermal extent of the atomic clouds along **y** before loading the optical lattice, the initial molecule distribution spanned ~12 lattice layers, with a peak of ~1500 molecules per layer. Compared with our previous work in two dimensions (*19*), where an auxiliary optical trap was used to compress KRb into few lattice layers to increase peak density, we deliberately prepared a broad distribution of molecules to minimize population differences among the central layers before layer addressing.

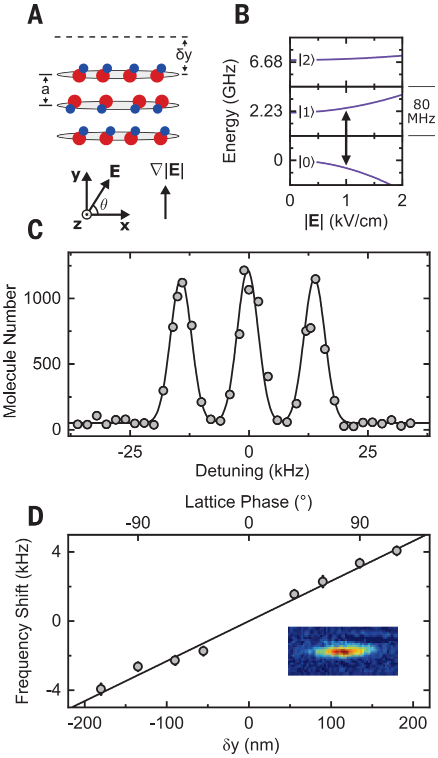


Fig. 1. Experimental configuration and individual layer addressing. (**A**) Molecules occupied 2D layers in the **x**-**z** plane, separated by layer spacing a. The bias electric field **E** was oriented at an angle θ in the **x**-**y** plane, with an electric field gradient $\nabla|\mathbf{E}|$ parallel to **y**. The lattice layers were displaced relative to the electrodes generating **E** by a distance δy. (**B**) KRb rotational structure. The arrow indicates the layer selection transition. (**C**) $|0\rangle \leftrightarrow |1\rangle$ frequency spectrum of a trilayer at $\partial_y|\mathbf{E}| = 6.4(2)\ \mathrm{kV\ cm^{-2}}$. Only three adjacent lattice layers were populated. (**D**) Center frequency shift of layer selection versus δy. Displacements smaller than 20 nm were measured. Error bars are 1 standard error from fits to the rotational transition line shape. (Inset) Absorption image of a single layer.

We generated highly configurable electric fields in the **x**-**y** plane using a set of six in-vacuum electrodes (*19*, *33*, *46*). To orient the induced dipole moments of the molecules, we rotated **E** by a variable angle θ, where θ = 0° corresponds to $\mathbf{E} \| \mathbf{x}$ (Fig. 1A). An electric field gradient $\nabla|\mathbf{E}|$ could be applied along **y**, parallel to the direction of tight confinement in the optical lattice. Because the KRb rotational energy levels are sensitive to $|\mathbf{E}|$, addressing single layers required stabilizing the molecule position relative to the local electric field distribution; such stabilization was accomplished by interferometrically measuring the positions of the lattice layers and electrodes and minimizing relative displacements (*45*). By adding a fixed phase offset to one of the beams generating the lattice, the layers were displaced from the electrodes along **y** by a distance δy. Using microwave pulses, the molecules could be transferred between states $|0\rangle$, $|1\rangle$, and $|2\rangle$ (Fig. 1B). Each rotational state $|N\rangle$ has an induced dipole moment d_N parallel to $\mathbf{E}$, the magnitude of which depends on $|\mathbf{E}|$, leading to a state-dependent energy shift of $-d_N|\mathbf{E}|$. At $|\mathbf{E}| = 1\ \mathrm{kV\ cm^{-1}}$, where all subsequent measurements were performed, the sensitivity of the $|0\rangle \leftrightarrow |1\rangle$ transition to $|\mathbf{E}|$ is 40 kHz $(\mathrm{V\ cm^{-1}})^{-1}$.

The electric field dependence of rotational state energies enabled microwave addressing of individual lattice layers. In terms of the layer spacing a and the dipole moments d_0 and d_1, a field gradient $\partial_y|\mathbf{E}|$ shifts the $|0\rangle \leftrightarrow |1\rangle$ transition on adjacent layers by the frequency

$$\Delta = \partial_y|\mathbf{E}| \cdot a \cdot \frac{d_0 - d_1}{h} \qquad (1)$$

With a microwave pulse of sufficiently narrow spectral width, all molecules in a single layer could be addressed without a measurable effect on other layers. In addition to layer-selective addressing of the $|0\rangle \leftrightarrow |1\rangle$ transition, we had the capability to apply global microwave pulses (addressing all molecules, irrespective of $\partial_y|\mathbf{E}|$) on the $|0\rangle \leftrightarrow |1\rangle$ and $|1\rangle \leftrightarrow |2\rangle$ transitions, as well as to globally remove $|0\rangle$ and $|1\rangle$ molecules with optical pulses of STIRAP light. From an initial condition of many occupied $|0\rangle$ layers, we used sequences of these microwave and optical pulses to prepare arbitrary layer configurations containing states $|0\rangle$, $|1\rangle$, and $|2\rangle$, including isolated 2D layers (*45*).

To demonstrate layer selection, we prepared three adjacent layers in $|1\rangle$ and scanned the frequency of an additional layer-selective $|0\rangle \leftrightarrow |1\rangle$ pulse in an applied gradient $\partial_y|\mathbf{E}| = 6.4(2)\ \mathrm{kV\ cm^{-2}}$. By monitoring the population transferred to $|0\rangle$ as a function of frequency, we probed the initial $|1\rangle$ distribution (Fig. 1C). We measured ~1200 molecules in each occupied layer, with adjacent layers detuned by 14 kHz. No molecules were measured in layers outside the trilayer or between occupied layers, which confirmed that the pulses were selectively addressing individual layers.

By varying the layer displacement δy and tracking the layer selection transition frequency, $|\mathbf{E}(y)|$ could be extracted with high spatial resolution, far below the interlayer spacing of 540 nm or the imaging diffraction limit. To characterize this technique, we probed an applied gradient of $\partial_y|\mathbf{E}| = 6.4(2)\ \mathrm{kV\ cm^{-2}}$. At each of eight different values of δy, spanning 360 nm (corresponding to a lattice phase shift of 240°), we measured the central frequency for layer selection (Fig. 1D). Fitting the frequency shift as a function of δy, we extracted $\partial_y|\mathbf{E}| = 5.8(3)\ \mathrm{kV\ cm^{-2}}$, with a maximum offset between δy and the line of best fit of only 20 nm. These measurements demonstrated subwavelength detection of molecule distributions using electric field gradients and high-precision electric field microscopy on nanometer spatial scales.

Using layer-selective addressing, we next optimized the rotational coherence in a single layer. Long-lived coherence is essential for realizing strong dipolar interactions (*9*, *14*). However, inhomogeneous broadening from external electric fields and optical trapping potentials tends to limit coherence. Because the molecules in two dimensions occupied a large spatial extent in the radial direction, electric field gradients transverse to **y** contribute a position-dependent frequency shift between rotational states (Fig. 1A). Spatial variation in the optical trap intensity contributes an additional frequency shift because of the differential ac polarizability between rotational states. The magnitude of the shift depends on $|\mathbf{E}|$ as well as the angle between the quantization axis (aligned with **E**) and the optical lattice polarization $\boldsymbol{\varepsilon}$ (*47*). In our apparatus, because the optical lattice polarization $\boldsymbol{\varepsilon}$ was fixed parallel to **x**, this angle coincided with the rotation angle θ of **E** (Fig. 1A). State-insensitive trapping can be achieved at the so-called magic angle of ~54°, where the differential ac polarizability between all rotational states with $m_N = 0$ vanishes (Fig. 2A).

We measured the Ramsey coherence of a single 2D layer with $\partial_y|\mathbf{E}| = 0$ at both θ = 90° and $\theta = \theta_m$, where we define θ_m as the electric field angle at which we measured the minimum differential polarizability (*45*). For both angles, the ODT polarization was set to the magic angle with respect to **E**. Using a single layer removed possible systematics, such as dipolar interactions between layers, stray electric field gradients along **y**, and layer-to-layer optical trap intensity variation. To measure the coherence decay, we prepared a single layer of molecules, used a $\pi/2$ pulse to initialize all molecules on the layer in an equal superposition of $|0\rangle$ and $|1\rangle$, held for a variable evolution time, applied a second $\pi/2$ pulse, and simultaneously measured the population in both states. As a function of evolution time *t*, we fit the contrast envelope to the Gaussian function e^{-t^2/τ^2}, where τ is the coherence time. For θ = 90°, we measured τ = 310(30) μs (Fig. 2B, top). This result was consistent with simulations where the differential ac polarizability was the only mechanism causing decoherence (*45*).

At $\theta = \theta_m$, little contrast decay was observed over the course of 600 μs (Fig. 2B, bottom). At longer evolution times, the Ramsey oscillation phase was scrambled by slight changes in $|\mathbf{E}|$ between experimental runs. We therefore computed the contrast on the basis of the observed variance of the rotational state populations (*45*). We measured τ = 1450(80) μs (Fig. 2C)—a factor of 5 improvement over θ = 90° and exceeding the longest bulk coherence time previously observed for KRb (*22*). Factors that may have limited the maximum achieved coherence included any remaining differential ac polarizability, residual electric field gradients, and intralayer dipolar interactions. With millisecond-scale coherence times and realistic experimental parameters, KRb is predicted to dynamically generate spin-squeezed states in two dimensions (*14*).

The capability to prepare arbitrary layer configurations enables the realization of new interacting systems. In this work, we studied a system where the rate of chemical reactions on a single 2D layer could be controlled by the presence of adjacent layers (*42*). The dynamics of molecules on multiple layers depend on a number of processes (Fig. 3A). In a single layer, molecules undergo two-body chemical reactions according to the rate equation $dN/dt = -\beta N^2$, where *N* is the molecule number and β is the two-body rate coefficient. KRb is fermionic, so ultracold molecules in the same internal state undergo reactions in the p-wave channel, with rate coefficient β_p. Molecules in different rotational states are distinguishable and therefore react in the s-wave channel, with rate coefficient β_s, which is typically orders of magnitude higher than β_p because of the absence of a centrifugal barrier (*24*, *48*). Molecules in separate layers in different rotational states may also exchange rotational angular momenta through long-range dipolar interactions (*41*), changing harmonic oscillator modes in the process (represented by γ in Fig. 3A). Spin exchange can only occur between states of opposite parity, meaning that $|0\rangle \leftrightarrow |1\rangle$ exchange is allowed and $|0\rangle \leftrightarrow |2\rangle$ exchange is forbidden at $|\mathbf{E}| = 0$. Applying an electric field induces rotational state mixing, but this effect only slightly weakens the above selection rules at $|\mathbf{E}| = 1\,\mathrm{kV\,cm^{-1}}$ (*45*).

Spin exchange facilitates the mixing of rotational state populations between initially spin-polarized layers, causing molecules undergoing exchange to be rapidly lost with rate coefficient β_s (*42*). To distinguish exchange from chemical reactions, we first measured the rate coefficients β_p and β_s at the temperature *T* = 334(30) nK, with $\nabla|\mathbf{E}| = 0$. Throughout the following, we describe bilayer and trilayer configurations according to the rotational states present in layers containing molecules. For example, 202 refers to a central layer containing only molecules of state $|0\rangle$ with adjacent $|2\rangle$ layers above and below and all other layers unoccupied. To extract β_p, which in general depends on rotational state because of variation

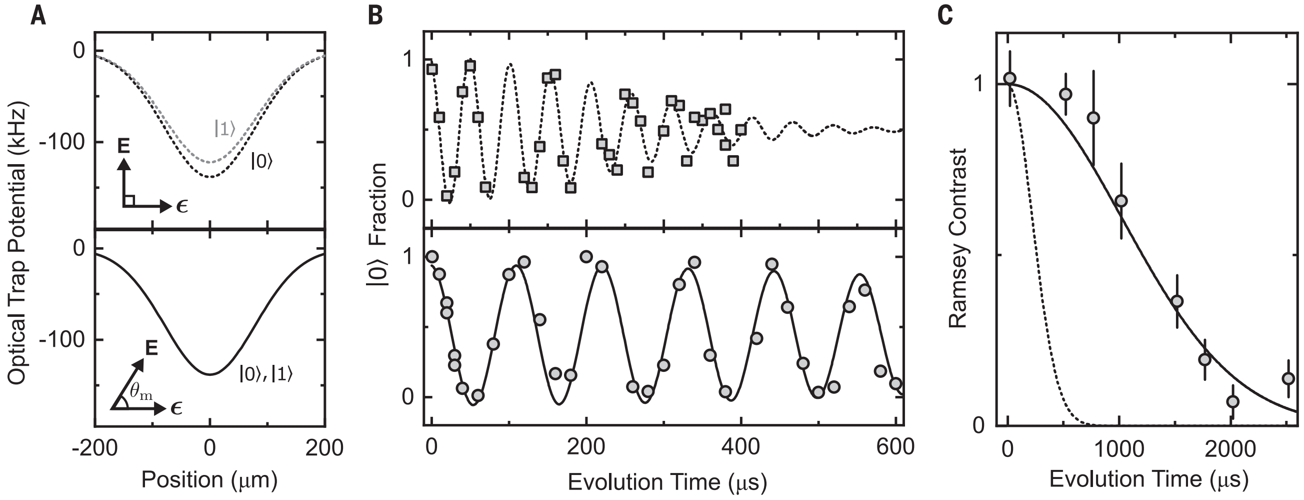


Fig. 2. Increasing rotational coherence time by rotating E. (**A**) Calculated optical trap potentials for states $|0\rangle$ and $|1\rangle$ in the ODT and optical lattice. The differential polarizability between rotational states depends on the angle θ between **E** and the optical lattice polarization $\boldsymbol{\varepsilon}$ and vanishes at $\theta_m \approx 54°$. (**B**) Ramsey oscillations of a single layer of molecules at θ = 90° (top, dashed line) and $\theta = \theta_m$ (bottom, solid line). (**C**) Contrast of Ramsey fringes at long evolution times. The coherence time at $\theta = \theta_m$ (points, solid line) was increased by a factor of 5 compared with θ = 90° (dashed line). Error bars are 1 standard error, estimated by simulations (*45*).

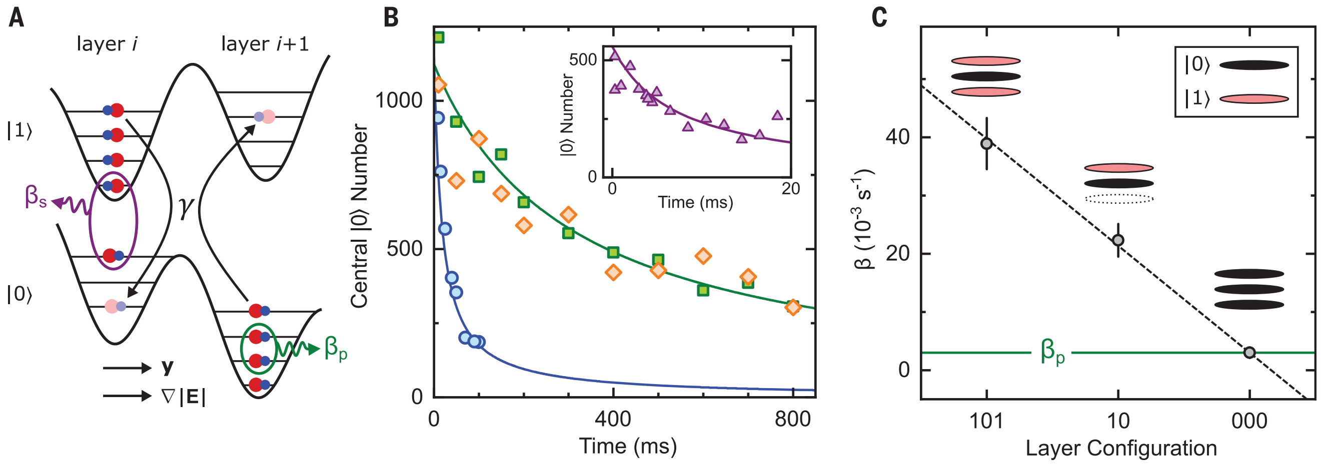


Fig. 3. Interaction and loss dynamics for molecules in two dimensions. (**A**) In a single layer, molecules in the same and in different rotational states undergo two-body loss with rate coefficients β_p and β_s, respectively. Molecules in different rotational states in adjacent layers may also exchange rotational states with rate γ, potentially changing harmonic oscillator modes during the exchange. (**B**) Central layer $|0\rangle$ molecule number versus time for 000 (β_p, green squares), 101 (blue circles), and 202 (orange diamonds) trilayers. The solid lines are fits to the two-body loss rate equation. (Inset) Loss for an equal mixture of molecules in states $|0\rangle$ and $|1\rangle$ (β_s). (**C**) Density dependence of spin exchange. β fit to the loss of $|0\rangle$ from layer configurations 000, 10, and 101 scaled linearly with the number of adjacent layers containing $|1\rangle$. The solid green line indicates β_p. Error bars are 1 standard error from fits to the two-body loss rate equation.

in the intermolecular potentials (*49*), we prepared spin-polarized 000 (Fig. 3B, green squares), 111, and 222 trilayers and fit the number decay to the two-body loss rate equation. For $|0\rangle$ and $|1\rangle$, $\beta_p = 2.99(17) \times 10^{-3}$ s^{-1}. For $|2\rangle$, the loss rate was reduced to $\beta_p = 1.78(24) \times 10^{-3}$ s^{-1}. To extract β_s, we prepared a 111 trilayer and applied a $\pi/2$ pulse to form an equal superposition of $|0\rangle$ and $|1\rangle$, which decohered completely within several milliseconds (Fig. 3B, inset). We measured $\beta_s = 2.0(3) \times 10^{-1}$ s^{-1}, nearly two orders of magnitude larger than β_p, as has been previously observed (*24*).

The interplay of exchange and loss was evident in layer configurations where multiple rotational states were present. In a 202 trilayer, where selection rules prohibit spin exchange, the loss rate of $|0\rangle$ molecules from the central layer matched β_p (Fig. 3B, orange diamonds). By contrast, for a 101 trilayer, the effective two-body loss rate increased by more than a factor of 10 (Fig. 3B, blue circles). The spin exchange rate depended on the density of molecules in adjacent layers, which is analogous to the dependence of the chemical reaction rate on the local molecule density. We demonstrated this effect by preparing 000, 10, and 101 layer configurations and fitting the decay of $|0\rangle$ molecules in the central layer to the two-body loss rate equation. The fit β scaled linearly with the number of adjacent $|1\rangle$ layers (Fig. 3C). These results showed the dependence of spin exchange on density and rotational state and demonstrated tuning of the chemical reaction rate using experimental control of the layer configuration.

To extract the spin exchange rate quantitatively, we described N_i^σ, the molecule number in layer i and in rotational state $|\sigma\rangle$, by a set of coupled differential equations including the aforementioned loss and exchange processes

$$\frac{dN_i^\sigma}{dt} = -\beta_p N_i^\sigma N_i^\sigma - \beta_s N_i^\sigma N_i^{\sigma'} + \gamma \sum_{k=i\pm1} \left(N_i^{\sigma'} N_k^\sigma - N_i^\sigma N_k^{\sigma'}\right) \quad (2)$$

where $\sigma \neq \sigma'$ are the two rotational states participating in the dynamics. The first two terms represent intralayer two-body loss, with rates β_p and β_s for spin-polarized and spin-mixed molecules, respectively. The third term represents spin exchange, which depends on the molecule populations in different rotational states in adjacent layers, occurring with rate constant γ (Fig. 3A). γ is an effective parameter describing the spin exchange, averaged over all molecules and over the full duration of the measurement.

Because spin exchange is a resonant process, adding an energy offset between adjacent layers suppresses its rate. To probe the energy spectrum of exchange, we added a variable gradient $\partial_y|\mathbf{E}|$ (Fig. 3A). The total change in electric potential energy when molecules in adjacent layers exchange rotational states is $h\Delta$ (Eq. 1), which is equivalent to the shift in microwave transition energy between adjacent layers (Fig. 1C). For states $|0\rangle$ and $|1\rangle$, Δ = 14 kHz at $\partial_y|\mathbf{E}| = 6.4$ kV cm^{-1}, the gradient used for layer selection.

We measured the spin exchange rate γ as a function of Δ in 101 and 202 trilayer configurations (Fig. 4A), with $\theta = 90°$ and at $T = 334(30)$ nK. For 202, the measured γ was consistent with zero spin exchange and did not depend on Δ. For 101, however, the peak exchange rate was $\gamma = 7.0(6) \times 10^{-3}$ s^{-1}, more than two times β_p. Notably, γ remained nonzero for large Δ, with a Lorentzian fit to $\gamma(\Delta)$ having a full width at half maximum (FWHM) of 6.4(6) kHz. This energy scale vastly exceeded the dipolar interaction energy between two molecules: At a separation of 540 nm, the rate of spin exchange between molecules in $|0\rangle$ and $|1\rangle$ is only 100 Hz (*41*).

Thermal energy contributed to the broad linewidth. To compensate Δ and conserve energy during spin exchange, molecules must change harmonic modes (Fig. 3A), the initial occupation of which is determined by the temperature. Qualitatively, this mechanism gives insight into the scaling of γ with T and Δ. At low temperatures, no spin exchange can occur when Δ greatly exceeds the thermal energy. For small Δ, however, the spin exchange rate is enhanced because of the high average occupation and strong dipolar coupling of low-lying harmonic modes. At high temperatures, the situation is reversed: High-lying modes are occupied, allowing exchange even at large Δ, but the peak exchange rate on resonance is suppressed. These effects suggest that increasing the temperature should broaden the spin exchange linewidth.

To quantify the temperature dependence, we repeated the measurement of $\gamma(\Delta)$ with a 101 trilayer at four temperatures between 370 and 650 nK (Fig. 4B). We set $\theta = \theta_m$ to eliminate possible broadening because of varying trap potentials between rotational states (Fig. 2A). At this angle, the strength of dipolar interactions between harmonic modes was slightly altered because the dipole moments were rotated relative to the plane of motion. At the

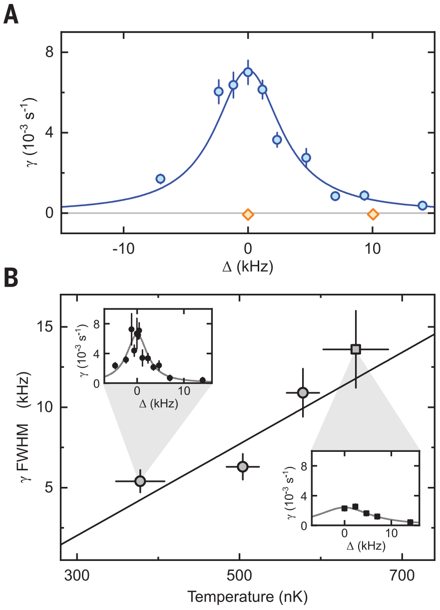


Fig. 4. Dependence of spin exchange on interlayer detuning Δ, rotational state, and temperature. (**A**) Spin exchange rate versus Δ, with θ = 90° and *T* = 334(30) nK, for 101 (blue circles) and 202 (orange diamonds) trilayers. The solid line is a Lorentzian fit to extract the FWHM. The point displayed at Δ = 0 kHz is the weighted average of measurements at Δ = 0 and ±0.12 kHz. (**B**) Temperature dependence of spin exchange linewidth, with $\theta = \theta_m$. The solid line is a linear fit for the temperature range shown. (Insets) γ versus Δ at *T* = 378(30) nK (circles, upper inset) and *T* = 643(40) nK (squares, lower inset). Error bars are 1 standard error from fits to Eq. 2.

lowest temperature, we measured a FWHM of 5.4(7) kHz, only slightly narrower than at θ = 90°. At the highest temperature, by contrast, the peak γ was reduced and the FWHM more than doubled to 13.6(2.4) kHz. To provide physical insight into the temperature dependence, we developed a simple two-particle model of molecules in adjacent layers interacting through dipolar spin exchange (*45*). Weighting by the thermal mode occupation, we found qualitative agreement with the observed trend. In addition to thermal energy, effects that may contribute to broadening include many-body interactions, where multiple molecules participate jointly in the spin exchange process, and intralayer dipole-dipole interactions. Both of these mechanisms should only weakly affect the temperature scaling because of the relatively low molecule density and small dipole moments at $|\mathbf{E}| = 1\,\mathrm{kV\,cm^{-1}}$.

We have demonstrated experimental control over spin exchange and chemical reactions in 2D systems of ultracold molecules, enabled by subwavelength addressing of individual lattice layers. These results provide a general method for layer-resolved state preparation and imaging of polar molecules, facilitating the study of many-body phases and nonequilibrium dynamics in long-range interacting systems with reduced dimensionality.

REFERENCES AND NOTES

1. A. Micheli, G. K. Brennen, P. Zoller, *Nat. Phys.* **2**, 341–347 (2006).
2. K. Góral, L. Santos, M. Lewenstein, *Phys. Rev. Lett.* **88**, 170406 (2002).
3. M. A. Baranov, M. Dalmonte, G. Pupillo, P. Zoller, *Chem. Rev.* **112**, 5012–5061 (2012).
4. J. L. Bohn, A. M. Rey, J. Ye, *Science* **357**, 1002–1010 (2017).
5. N. Y. Yao *et al.*, *Phys. Rev. Lett.* **110**, 185302 (2013).
6. A. V. Gorshkov *et al.*, *Phys. Rev. A* **84**, 033619 (2011).
7. S. R. Manmana, E. M. Stoudenmire, K. R. A. Hazzard, A. M. Rey, A. V. Gorshkov, *Phys. Rev. B* **87**, 081106 (2013).
8. N. Y. Yao, M. P. Zaletel, D. M. Stamper-Kurn, A. Vishwanath, *Nat. Phys.* **14**, 405–410 (2018).
9. D. Peter, S. Müller, S. Wessel, H. P. Büchler, *Phys. Rev. Lett.* **109**, 025303 (2012).
10. S. V. Syzranov, M. L. Wall, V. Gurarie, A. M. Rey, *Nat. Commun.* **5**, 5391 (2014).
11. N. R. Cooper, G. V. Shlyapnikov, *Phys. Rev. Lett.* **103**, 155302 (2009).
12. J. Levinsen, N. R. Cooper, G. V. Shlyapnikov, *Phys. Rev. A* **84**, 013603 (2011).
13. M. Babadi, E. Demler, *Phys. Rev. A* **86**, 063638 (2012).
14. T. Bilitewski *et al.*, *Phys. Rev. Lett.* **126**, 113401 (2021).
15. D.-W. Wang, M. D. Lukin, E. Demler, *Phys. Rev. Lett.* **97**, 180413 (2006).
16. A. C. Potter, E. Berg, D.-W. Wang, B. I. Halperin, E. Demler, *Phys. Rev. Lett.* **105**, 220406 (2010).
17. A. Pikovski, M. Klawunn, G. V. Shlyapnikov, L. Santos, *Phys. Rev. Lett.* **105**, 215302 (2010).
18. N. T. Zinner, B. Wunsch, D. Pekker, D.-W. Wang, *Phys. Rev. A* **85**, 013603 (2012).
19. G. Valtolina *et al.*, *Nature* **588**, 239–243 (2020).
20. Z. Zhang, L. Chen, K.-X. Yao, C. Chin, *Nature* **592**, 708–711 (2021).
21. J. S. Rosenberg, L. Christakis, E. Guardado-Sanchez, Z. Z. Yan, W. S. Bakr, Observation of the Hanbury Brown and Twiss effect with ultracold molecules. arXiv:2111.09426 [cond-mat.quant-gas] (2021).
22. B. Neyenhuis *et al.*, *Phys. Rev. Lett.* **109**, 230403 (2012).
23. F. Seeßelberg *et al.*, *Phys. Rev. Lett.* **121**, 253401 (2018).
24. S. Ospelkaus *et al.*, *Science* **327**, 853–857 (2010).
25. M.-G. Hu *et al.*, *Science* **366**, 1111–1115 (2019).
26. R. Bause *et al.*, *Phys. Rev. Res.* **3**, 033013 (2021).
27. P. Gersema *et al.*, *Phys. Rev. Lett.* **127**, 163401 (2021).
28. P. D. Gregory, J. A. Blackmore, S. L. Bromley, S. L. Cornish, *Phys. Rev. Lett.* **124**, 163402 (2020).
29. G. Quéméner, J. L. Bohn, *Phys. Rev. A* **81**, 060701 (2010).
30. K.-K. Ni *et al.*, *Nature* **464**, 1324–1328 (2010).
31. M. H. G. de Miranda *et al.*, *Nat. Phys.* **7**, 502–507 (2011).
32. A. Frisch *et al.*, *Phys. Rev. Lett.* **115**, 203201 (2015).
33. K. Matsuda *et al.*, *Science* **370**, 1324–1327 (2020).
34. J.-R. Li *et al.*, *Nat. Phys.* **17**, 1144–1148 (2021).
35. L. Anderegg *et al.*, *Science* **373**, 779–782 (2021).
36. G. Wang, G. Quéméner, *New J. Phys.* **17**, 035015 (2015).
37. T. Karman, J. M. Hutson, *Phys. Rev. Lett.* **121**, 163401 (2018).
38. M. Karski *et al.*, *New J. Phys.* **12**, 065027 (2010).
39. J. F. Sherson *et al.*, *Nature* **467**, 68–72 (2010).
40. G. J. A. Edge *et al.*, *Phys. Rev. A* **92**, 063406 (2015).
41. B. Yan *et al.*, *Nature* **501**, 521–525 (2013).
42. A. Pikovski, M. Klawunn, A. Recati, L. Santos, *Phys. Rev. A* **84**, 061605 (2011).
43. K.-K. Ni *et al.*, *Science* **322**, 231–235 (2008).
44. L. De Marco *et al.*, *Science* **363**, 853–856 (2019).
45. See the supplementary materials.
46. J. Covey, thesis, University of Colorado Boulder (2017).
47. S. Kotochigova, D. DeMille, *Phys. Rev. A* **82**, 063421 (2010).
48. Z. Idziaszek, P. S. Julienne, *Phys. Rev. Lett.* **104**, 113202 (2010).
49. S. Kotochigova, *New J. Phys.* **12**, 073041 (2010).
50. W. G. Tobias *et al.*, Data for "Reactions between layer-resolved molecules mediated by dipolar spin exchange", version 1, Zenodo (2022); https://doi.org/10.5281/zenodo.6081435.

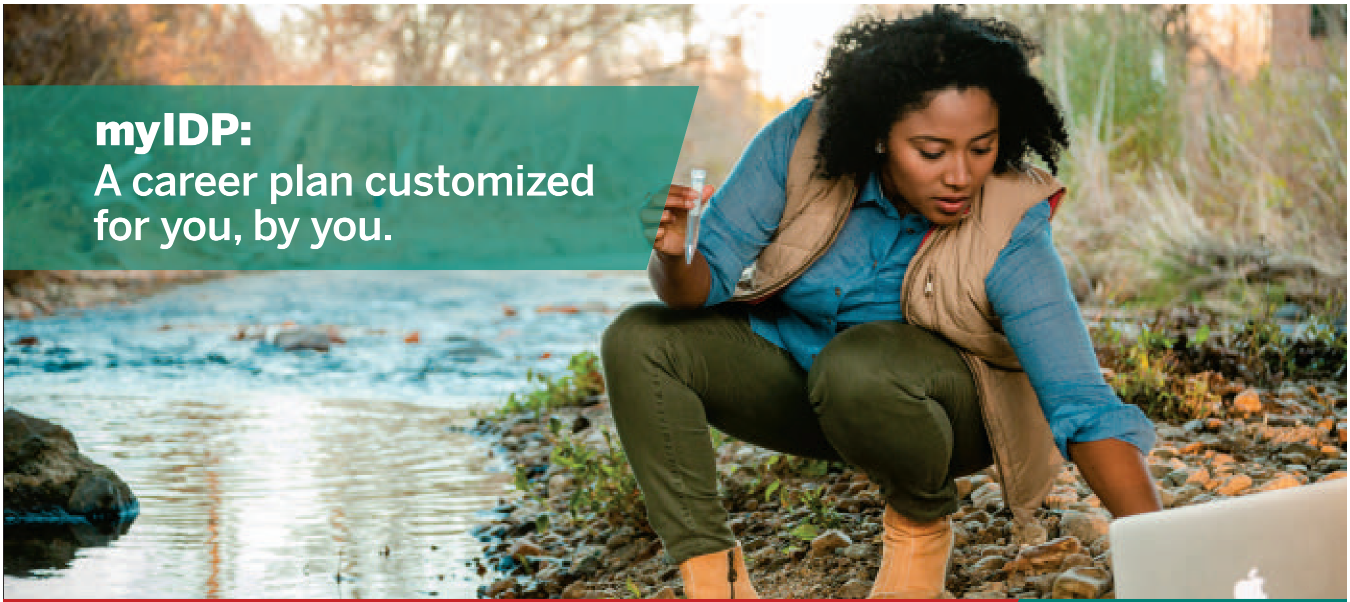
ACKNOWLEDGMENTS

We thank S. R. Cohen, L. De Marco, and G. Valtolina for experimental contributions and Y. Liu and C. Sanner for helpful discussions. **Funding:** This study received support from the Army Research Office, Multidisciplinary University Research Initiative; the Air Force Office of Scientific Research, Multidisciplinary University Research Initiative; the Defense Advanced Research Projects Agency, Driven and Nonequilibrium Quantum Systems; the National Science Foundation, Quantum Leap Challenge Institutes, Office of Multidisciplinary Activities (grant 2016244); the National Science Foundation (grant Phys-1734006); and the National Institute of Standards and Technology. **Author contributions:** Experimental work and data analysis: W.G.T., K.M., J.-R.L., C.M., A.N.C., and J.Y. Two-particle exchange model: T.B. and A.M.R. Interpreting results and writing the manuscript: W.G.T., K.M., J.-R.L., C.M., A.N.C., T.B., A.M.R., and J.Y. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in this paper are present in the paper or the supplementary materials and are available through Zenodo (*50*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn8525
Supplementary Text
Figs. S1 to S4

24 December 2021; accepted 15 February 2022
10.1126/science.abn8525



UC DAVIS HEALTH

The Department of Medical Microbiology and Immunology in the School of Medicine at the University of California, Davis, announces a search for a qualified candidate with diagnostic and administrative capabilities to direct the diagnostic activities of a Fungal Clinical Diagnostic Laboratory (currently emphasizes coccidioidomycosis serology). The candidate will be recruited for a full-time position at the **Assistant**, **Associate** or **Full Professor** rank in the In Residence series. Appointments to this series are expected to engage in responsibilities that include provision of serologic tests for coccidioidomycosis, reviewing results of serologic tests, providing consultation and managing laboratory staff.

This position offers opportunities for biomedical investigations. Faculty in this series are expected to engage in teaching, independent research and creative work, demonstrate professional competence and activity, and provide University and public service.

Candidates must possess a Ph.D. and be board certified with the American Board of Medical Laboratory Immunology, or the American Board of Medical Microbiology, or be board certified in Pathology or any other boards approved under the Clinical Laboratory Improvement Act. Or possess an M.D., be board certified in Internal Medicine and Infectious Diseases or Pathology. A background in teaching and administration is desirable. Collaborations with clinical and research programs at UC Davis and the UC Davis Medical Center are available. The successful candidate must be able to demonstrate that they are legally authorized to work in the United States. The University will not offer sponsorship for this position.

For full consideration, applications should be received by **April 18, 2022**; however, the position will remain open until filled through **June 30, 2022**. Qualified applicants should upload a cover letter, curriculum vitae, statement of teaching, statement of research, statement of contributions to diversity and 3-5 letters of recommendation online at **https://recruit.ucdavis.edu/apply/JPF04764**

Science Careers helps you advance your career. Learn how !

- Register for a free online account on **ScienceCareers.org**.
- Search hundreds of job postings and find your perfect job.
- Sign up to receive e-mail alerts about job postings that match your criteria.
- Upload your resume into our database and connect with employers.
- Watch one of our many webinars on different career topics such as job searching, networking, and more.

Visit **ScienceCareers.org**
today — all resources are free

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

SCIENCECAREERS.ORG

上海交通大学 SHANGHAI JIAO TONG UNIVERSITY 生物医学工程学院 School of Biomedical Engineering

Dean of the School of Biomedical Engineering
Shanghai Jiao Tong University
Shanghai, China

We invite nominations and applications for the position of Dean of the School of Biomedical Engineering (BME) at Shanghai Jiao Tong University (SJTU). The School of BME at SJTU was founded in 2011, but the history of our BME program can be dated back to 1979, when the major of biomedical instrumentation was first established. Our School is the first in China to have its BME program accredited by ABET. It has been consistently ranked among the top three in China by the Ministry of Education, based on national assessment conducted every five years.

Our undergraduate and graduate education programs encompass three major directions: 1) biomedical instrumentation; 2) nano, molecular and regenerative medicine; and 3) imaging, computational and systems biomedicine. The School has about 100 full time faculty members and nearly 900 full time students, including over 600 graduate students. To promote interdisciplinary research and clinical translation, the School has led a number of well recognized collaborative centers of excellence, including the Med-X Research Institute, established to conduct clinically driven basic research and development, supported by joint research centers with top ranked hospitals in China.

We expect that the new Dean will bring exceptional leadership skills, energy and enthusiasm to bring the School to a new height. The candidate should have extensive experience in higher education and academic administration, a strong and innovative vision, and an established track record in multidisciplinary research and international collaboration.

Position profile

The Dean is the chief academic and administrative officer of the School, expected to set the strategic vision and provide leadership in the School's pursuit of its goals of providing high quality undergraduate and graduate education and research. Key to the responsibilities of this position, this leader will:

- pursue innovative and strategic research directions;
- recruit and maintain an excellent faculty;
- create and support high quality educational programs;
- attract excellent undergraduate and graduate students;
- expand interdisciplinary research engagements and impacts nationally/internationally;
- enhance productive relationships with alumni, industries, and local, national, and international communities.

Candidate Qualifications

An earned doctorate, at the full professor level, a record of progressive and visionary academic leadership experience in higher education (i.e. department chair or research center director, etc.) are desired. The ideal candidate should be an internationally recognized scholar with the desire, ability, and vision to sustain a leadership role within an organization that is in the forefront of developing new models for research and scholarship, undergraduate and graduate education.

Applications should include a CV and a cover letter that articulates the candidate's interest and experience for the deanship at the School. These be emailed as a PDF document to **sjtuzzb@sjtu.edu.cn** before April 30 of 2022. Interested candidates could also send email to the above for any question. For more information, just check our website: http://en.bme.sjtu.edu.cn/show-59-413.html

WORKING LIFE

By **Nkuchia M. M’ikanatha**

The trouble with labels

A few years ago, I sat down to read a paper in *Science*. I was fascinated by the findings because they were relevant to my research in public health. But I also felt a sting of annoyance. The researchers stated they’d done their work in “developing countries”—a hierarchical term that, to me, is a slight to many parts of the world, including my native country, Kenya. It wasn’t an isolated incident: I repeatedly come across this kind of label in the academic literature and in discussions about global challenges such as the COVID-19 pandemic. And I’ve learned to speak up.

Three decades ago, I moved to the United States for graduate school. During my first semester, I went to the library to look at journal articles for a class assignment. I repeatedly encountered terms such as “third world,” “developing,” and “undeveloped,” used to describe countries like Kenya. Seeing those terms nauseated me. “Third world”—as if to label my people at the bottom of the world’s pecking order?

“The terms are … needlessly insulting to the people who live in those countries.”

Eventually, I decided I had to say something. I wrote a letter to the local newspaper. It was cathartic to get things off my chest, but I never expected it to be published.

A week later, I began to sweat when my adviser called me into his office and handed me a newspaper clipping. He remained quiet as I skimmed the article I quickly recognized as my own. Then he shared his thoughts. “You make important points. We *expect* you to question what you hear or read,” he said. It was a relief to be commended rather than chastised.

For the remainder of my studies, I urged my colleagues to think carefully about these terms. I told them that grouping countries together with one label often obscures important variation in how those countries allocate resources. Some African countries, for instance, invest much more in public health than others. The terms are also needlessly insulting to the people who live in those countries.

My fellow students were generally indifferent when I made these points. One, also from Africa, thought I was making too much trouble. “Our goal is to finish class and get our degrees, not to poke holes in terms that have already been decided,” he told me. But several professors appreciated my perspective. I felt that I might be making a difference.

After I graduated, I moved on to government public health work and felt less comfortable speaking up. I got the impression that my colleagues and supervisors wouldn’t welcome the challenges. Plus, as an immigrant working in the United States, I was self-conscious about how my behavior might be perceived. So I often remained quiet.

That began to change when I coedited a textbook with three thoughtful colleagues. I expressed my discomfort with the standard terms and proposed as an alternative “areas with limited resources.” The others agreed, so we included it in our guidelines for contributors. I was glad the resulting textbook was devoid of terms I dislike. But more recently, I have realized that even my own term is simplistic and rarely justifiable, and that it is often better to avoid these categorizations altogether.

Over the years, I have been heartened to see some progress in the broader publishing landscape. These days I rarely encounter “third world,” for instance. But “developing” is still used by public health experts—for example, in communications from the World Health Organization and in research articles, even those published by scientific societies that have issued strong social justice and diversity statements.

I encourage my colleagues to avoid using the term altogether. “Low income” is one alternative, but it’s often not relevant unless the research is strictly economic. Instead, I’d recommend sticking with the names of the actual locations. When discussing vaccine hesitancy in Africa compared to Europe, for instance, the term “developing” is irrelevant—simply use the names of the continents instead.

In the end, it’s important to remember that people live in the countries you are describing. Avoid demeaning language—because if you don’t you may inadvertently harm readers and perpetuate prejudice. ■

Nkuchia M. M’ikanatha is an epidemiologist in Pennsylvania and an adjunct scholar at the Perelman School of Medicine. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER