What's in store in 2020? *Science* looks ahead p. 6

A sci-fi legend's legacy lives on p. 20

Earliest cooked root vegetables p. 87

Science

$15
3 JANUARY 2020
sciencemag.org

AAAS

TINY CARNIVOROUS PLANTS

Convoluted shapes from simple rules pp. 24 & 91

CONTENTS

71
What makes knots tight?

3 JANUARY 2020 • VOLUME 367 • ISSUE 6473

9
Afterglow of a 1962 nuclear test in space

CREDITS (FROM TOP): (IMAGE) V. P. PATIL *ET AL.*; (PHOTO) SCIENCE HISTORY IMAGES/AL AMY STOCK PHOTO; (PHOTO) TONY HEALD/MINDEN PICTURES

NEWS

IN BRIEF

6 What's coming up in 2020
PODCAST

IN DEPTH

9 U.S. military tests radiation belt cleanup in space
Radio waves could sweep belts clean of satellite-killing particles after nuclear sneak attack *By R. Stone*

10 Past megadroughts hit North and South America in tandem
Strong La Niña conditions drove deep medieval droughts *By P. Voosen*

11 Study pushes emergence of measles back to antiquity
The virus may have entered the human population when cities grew large enough to sustain outbreaks *By K. Kupferschmidt*

12 Computer scientist in line to become next NSF director
Sethuraman Panchanathan would succeed France Córdova *By J. Mervis*

13 Congress again rejects Trump cuts, smiles on science agencies
2020 spending bill increases research budgets, but next year will be tougher as lawmakers will have less to spend
By J. Merivs

14 Global polio eradication falters in the final stretch
Vaccine-derived outbreaks may force a change in "endgame" strategy *By L. Roberts*

FEATURES

16 Fighting words
Virologist Roberto Burioni has become a celebrity in Italy by sparring with vaccine skeptics *By D. Starr*

6

INSIGHTS

BOOKS *ET AL.*

20 Asimov at 100
From epic space operas to rules for robots, the prolific author's literary legacy endures *By J. Gunn*

PERSPECTIVES

22 Uncovering the ART of antimalarial resistance
A key mechanism of resistance to the antimalarial drug artemisinin is identified
By D. Marapana and A. F. Cowman
RESEARCH ARTICLE p. 51

23 Majorana fermions go for a ride
Evidence for propagating Majorana quasiparticles is found in a topological superconductor
By S. Tewari and T. D. Stanescu
REPORT p. 104

24 Building a carnivorous trap
Experiments and computations reveal developmental origins of cup-shaped leaves
By D. E. Moulton and A. Goriely
REPORT p. 91

Published by AAAS

26 Electrostatics affect the glow
Chromophore twisting is probed with unnatural amino acids *By C. Hu* et al.
REPORT p. 76

27 Cancer in sub-Saharan Africa
Knowledge of cancer in Africa brings needed diversity to improve health worldwide
By T. R. Rebbeck

29 Gene expression regulated by RNA stability
The factor responsible for autoregulation of tubulin RNA stability is identified
By O. Shoshani and D. W. Cleveland
REPORT p. 100

POLICY FORUM

30 Sustainable minerals and metals for a low-carbon future
Policy coordination is needed for global supply chains *By B. K. Sovacool* et al.

LETTERS

34 NextGen Voices: Making science accessible

RESEARCH

IN BRIEF

36 From *Science* and other journals

REVIEW

39 Neuroscience
Memory engrams: Recalling the past and imagining the future
S. A. Josselyn and S. Tonegawa
REVIEW SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.AAW4325

RESEARCH ARTICLES

40 Nanoparticles
Oriented attachment induces fivefold twins by forming and decomposing high-energy grain boundaries *M. Song* et al.

45 Multiplex genomics
Massively multiplex chemical transcriptomics at single-cell resolution *S. R. Srivatsan* et al.

51 Malaria
A Kelch13-defined endocytosis pathway mediates artemisinin resistance in malaria parasites *J. Birnbaum* et al.
PERSPECTIVE p. 22

REPORTS

59 Topological optics
A single photonic cavity with two independent physical synthetic dimensions
A. Dutt et al.

64 Topological matter
Absence of evidence for chiral Majorana modes in quantum anomalous Hall-superconductor devices *M. Kayyalha* et al.

68 Superconductivity
Atomic manipulation of the gap in $Bi_2Sr_2CaCu_2O_{8+x}$ *F. Massee* et al.

71 Applied physics
Topological mechanics of knots and tangles *V. P. Patil* et al.

76 Fluorescent proteins
Electrostatic control of photoisomerization pathways in proteins *M. G. Romei* et al.
PERSPECTIVE p. 26

79 Optics
On-chip integrated laser-driven particle accelerator *N. V. Sapra* et al.

83 Neuroscience
Dendritic action potentials and computation in human layer 2/3 cortical neurons
A. Gidon et al.

87 Archaeology
Cooked starchy rhizomes in Africa 170 thousand years ago *L. Wadley* et al.

91 Plant science
Evolution of carnivorous traps from planar leaves through simple shifts in gene expression *C. D. Whitewoods* et al.
PERSPECTIVE p. 24; PODCAST

96 Evolutionary biology
Protein-coding changes preceded cis-regulatory gains in a newly evolved transcription circuit *C. S. Britton* et al.

100 Tubulin
TTC5 mediates autoregulation of tubulin via mRNA degradation *Z. Lin* et al.
PERSPECTIVE p. 29

104 Topological matter
Evidence for dispersing 1D Majorana channels in an iron-based superconductor *Z. Wang* et al.
PERSPECTIVE p. 23

DEPARTMENTS

5 Editorial
Clarity in 2020 *By H. Holden Thorp*

114 Working Life
Strange dreams *By Desiree Dickerson*

ON THE COVER

Immature trap of the carnivorous plant *Utricularia gibba* (~200 µm in width), illustrating its two-layered cellular structure. Such cup-shaped leaves have evolved multiple times as a mechanism for trapping animals. Analyzing *U. gibba*, researchers used a combination of molecular genetics and computer modeling to reveal how shifts in gene expression domains can account for repeated evolution of cup shapes from species with planar leaves. See pages 24 and 91. *Image: Karen Lee and Claire Bushell*

New Products.....110
Science Careers.....111

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 © 2020 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): $2148; 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.sciencemag.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

PHOTO: MICHAEL ROBINSON CHAVEZ/THE WASHINGTON POST VIA GETTY IMAGES

Published by AAAS

EDITORIAL

Clarity in 2020

We begin 2020 (or should I say 20/20) awash in references to ophthalmology. In that vein, we hope to make science clearer and brighter in the months ahead—which brings me to the issue of transparency.

The *Science* family of journals looks forward to a year in which the editors will strive for greater transparency and reproducibility in the science that we publish. Policies on conflicts of interest and professional behavior for authors are now uniform across our journals and strengthen our standards for disclosure. For the first time, we will require general disclosures from *Science*'s Board of Reviewing Editors; prior to this year, they were only asked to disclose conflicts that arose for papers on which they were consulted.

As readers of *Science*'s news section would know, I have been active for decades in the biotechnology business and disclosed competing interests to the American Association for the Advancement of Science (AAAS, publisher of the *Science* journals) upon joining the editorial team. These interests are stated on my web bio (www.sciencemag.org/about/leadership-and-management#HoldenThorp), and I commit to keeping them updated for full public disclosure. The *Science* journals believe that the commercial application of scientific findings is critical to gaining public support and to ensuring that scientists are involved in the realization of their ideas. At the same time, vigorous disclosure of these competing interests is a must.

"...editors will strive for greater transparency and reproducibility..."

We will also continue to strive for greater reproducibility in the science that we publish. Last year, *Science* helped to develop a framework setting out minimal expectations for materials, design, analysis, and reporting (MDAR) and also participated with other publishers in piloting a checklist that operationalized this framework (https://cos.io/blog/journals-test-materials-design-analysis-reporting-mdar-checklist/). Authors were admirably enthusiastic about this process, and we are pleased to endorse the framework and implement the checklist this year for life science papers in *Science*. This should allow information to be used by others seeking to reproduce findings and will hopefully pave the way for more such standardization.

One of the questions I am often asked is how the *Science* journals correct or retract papers. With the rapid analysis that can happen on social media, these questions are being raised faster than ever before. We don't seek to hide from these efforts: Image sleuth Elisabeth Bik (@MicrobiomDigest) recently stated on Twitter that she needed a list of journal editors' emails and I immediately responded with mine (which is not hard to find at right).

Science journals prefer to publish a retraction signed by all authors. Deciding to retract in cases where not all authors agree can require a lengthy investigation by universities. In such cases, we publish an Editorial Expression of Concern (EEoC) in the meantime, to quickly alert readers that concerns have been raised about the reported work. We hope that making this information available promptly helps mitigate the waste of time, effort, and funds by researchers who might otherwise base future work on papers that are later retracted. During my short time as editor-in-chief, when we have contacted universities to begin this process, I have been giving a deadline by which we will decide if we are going to proceed with an EEoC. After running academic units for the past 14 years, I know that it sometimes takes a nudge to get things to the top of an inbox. We also publish an EEoC when authors alert us to problems with their published paper and need time to determine whether the findings hold. Authors taking the initiative to correct the record strengthens the integrity of the scientific enterprise.

The year 2020 brings a presidential election to the United States and a time of transition across the Atlantic. There will be great opportunities for everyone to stand up for science and for their beliefs. Breakthroughs in all fields will continue with new ideas, insights, and applications, hopefully for the good of all members of society and for the Earth that we cherish. And it will be exciting for AAAS, as our new chief executive officer, Sudip Parikh, begins next week. I have known Sudip for 25 years and look forward to his generous and thoughtful leadership.

–H. Holden Thorp

PHOTO: CAMERON DAVIDSON

H. Holden Thorp
Editor-in-Chief, *Science* journals.
hthorp@aaas.org;
@hholdenthorp

10.1126/science.aba6293

Published by AAAS

NEWS

IN BRIEF

Edited by **Jeffrey Brainard**

AREAS TO WATCH

What's coming up in 2020

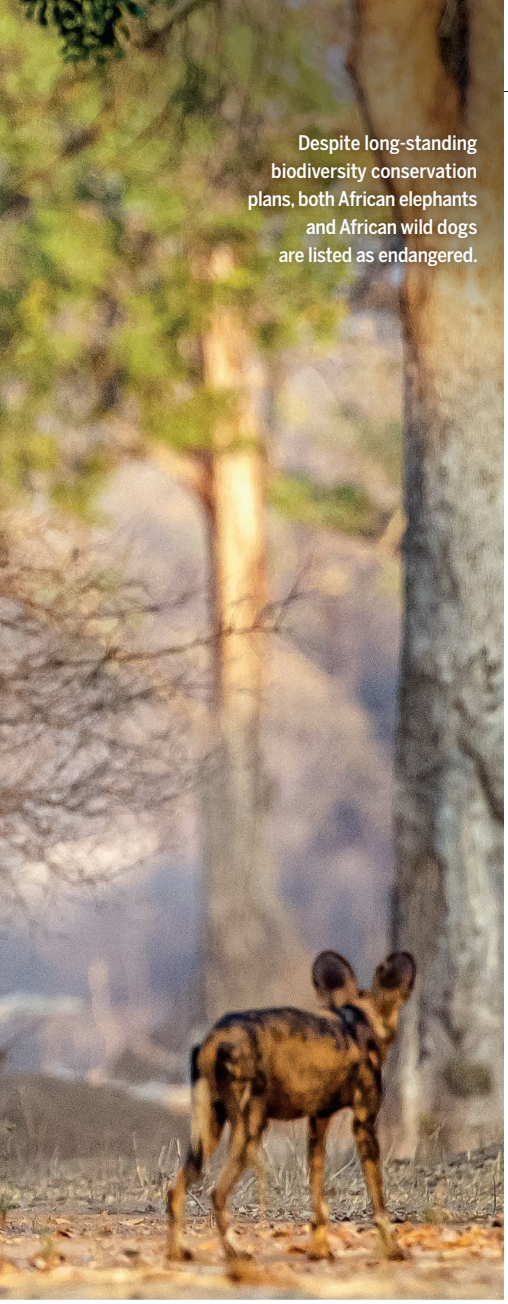
Incessant political turmoil in the United Kingdom, United States, and other nations will likely last well into the new year, complicating many researchers' work. The U.K. election last month made the country's departure from the European Union a near-certainty, and its scientists now face losing EU science grants and scientific collaborators. In the United States, a presidential election in November will determine the role of scientists in future policy deliberations; many experts on climate change and other environmental issues assert that the Trump administration has ignored scientific evidence. In this section, *Science*'s news staff forecasts other areas of policy and research likely to make news this year amid the chaos, from dark matter detectors to new efforts to rein in loss of species.

New goals for saving biodiversity

CONSERVATION | This year will see an attempt to revitalize the ambitious Aichi Biodiversity Targets, named for the city in Japan where they were negotiated. Since they were approved 10 years ago, there has been little to no progress in meeting most of those 20 goals, such as preventing the decline of endangered species. That alarming situation was highlighted last year in a major scientific assessment by another organization, the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. But in October, nations will have a chance to try to set a more effective course when they meet in Kunming, China, to review and revise the Convention on Biological Diversity, the world's flagship conservation pact.

PHOTO: TONY HEALD/MINDEN PICTURES

Published by AAAS



Despite long-standing biodiversity conservation plans, both African elephants and African wild dogs are listed as endangered.

PHOTO: K. MCGRATH *ET AL.*, *SCIENTIFIC REPORTS* **9**, 11027 (2019)

Crunch time for climate policy

POLICY | The politics of climate change faces crucial moments this year. The Trump administration's opposition to regulations reducing fossil fuel emissions has emerged as a primary talking point for the president's Democratic challengers. One day after the U.S. presidential election on 3 November, the country, the second largest emitter of greenhouse gases, is set to leave the Paris climate accord, although a Democratic president could quickly rejoin after taking office in 2021. Less than 1 week later, the United Nations will convene in Glasgow, U.K., for its most important climate summit since 2015, where nations are expected to increase their pledges to cut greenhouse gas emissions—even though they are behind on meeting existing ones. Without stepped-up efforts, there is little hope the world can keep future warming below 2°C, the level scientists forecast will produce catastrophic damage to human communities and ecosystems.

Counting on the census

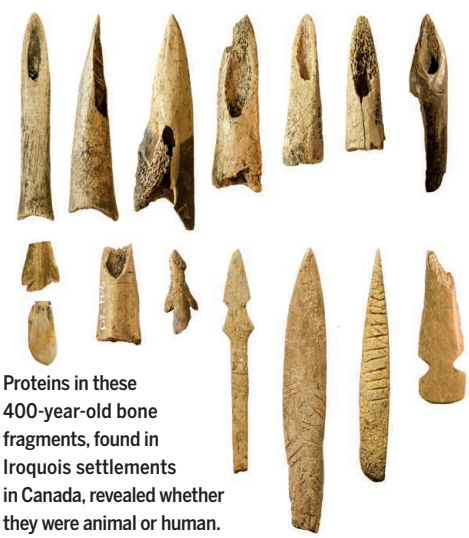
DEMOGRAPHICS | The U.S. government has conducted a decennial census since 1790. But the 2020 census that kicks off on 1 April faces unprecedented political challenges. Although civil rights groups won a fight to block a question about citizenship that the Trump administration wanted to add, it has ordered the Census Bureau to generate the equivalent data using existing government records so that states can use the information when they redraw boundaries for federal and state elections. Researchers fear that assignment may not be doable, and that the political debate has further alienated those already hardest to count. Demographers also worry that the census' use of a new way to protect respondents' privacy could distort analyses of demographic trends. At stake are not only how more than $1.5 trillion in federal funds are distributed each year, but also the integrity of the nation's largest statistical agency.

CRISPR's big clinical tests

BIOMEDICINE | The CRISPR gene-editing tool faces key tests this year of its promise to treat cancer and genetic diseases. A small U.S. clinical trial is using CRISPR to disable three genes in T cells that are then returned to a cancer patient's body, an approach that could help these immune system soldiers stop malignant cells from growing and extend patients' lives. More results may also come from separate CRISPR cancer trials in China. Other researchers are working to treat people with sickle cell disorder and thalassemia by using the DNA editor to turn on the gene for a fetal version of hemoglobin to compensate for a defective adult form of the oxygen-carrying protein; last fall, scientists reported success in two patients and in 2020 will present longer-term results for a larger group. Another clinical trial in the United States could show whether CRISPR improves vision in people with an inherited disorder that causes progressive blindness.

Proteins tell ancient tales

ARCHAEOLOGY | Ancient proteins will shed new light this year on the identity and behavior of humans and other animals that lived more than 1 million years ago. Proteins are more stable than DNA, and as analytical methods improve, researchers can apply them to understand more about older fossils lacking DNA, including the sex and age of remains of enigmatic ancient hominins. Most hominins are known by bones and teeth alone, and proteins could provide a new tool for sorting them in family trees and to identify fragments too small to classify. Although tooth enamel offers the best source of ancient proteins, researchers are also extracting them from bones and hair. In addition, proteins can reveal new information about artifacts made of plant and animal materials, and researchers hope this year to analyze parchment manuscripts and the beeswax once used to seal documents. Scientists are also analyzing residues on pots for more clues to whether early pastoralists in the steppelands of Mongolia, for example, drank camel or goat milk first—and what people living on the edge of the Roman Empire in England ate.



Proteins in these 400-year-old bone fragments, found in Iroquois settlements in Canada, revealed whether they were animal or human.

'Foreign influence' worries grow

NATIONAL SECURITY | The political debate over how to respond to China's emergence as a scientific superpower is likely to intensify this year. In the United States, some federal agencies have banned their employees from participating in foreign talent recruitment programs—an approach that China has used to connect with thousands of scientists—to prevent disclosure of information that could damage national security and U.S. economic competitiveness. Two new bodies created by Congress will work to harmonize practices across federal agencies and chew over how best to balance openness and security. U.S. academic leaders are hoping to convince policymakers not to fence off certain types of research, which they say would throttle U.S. innovation. A new report to the National Science

Published by AAAS

Foundation says teaching students and faculty members about acceptable and unacceptable behavior is a better approach.

Japan boosts neutrino efforts

PARTICLE PHYSICS | Japan is expanding neutrino research to better understand properties of the phantom particles and the cosmic processes that produce them. This spring, scientists will increase the sensitivity of the 22-year-old Super-Kamiokande neutrino observatory by doping water in its observation chamber with the rare-earth metal gadolinium. The detector will then watch for signals generated when neutrinos from supernovae hit the water, providing clues about the dynamics within those exploding stars. Japan's legislature is expected to fund an even bigger step: construction of the 72 billion Japanese yen ($660 million) Hyper-Kamiokande. Ten times larger than its predecessor, it will capture that much more data about neutrinos emanating from the Sun, distant stars, and supernovae.

Dueling dark matter detectors

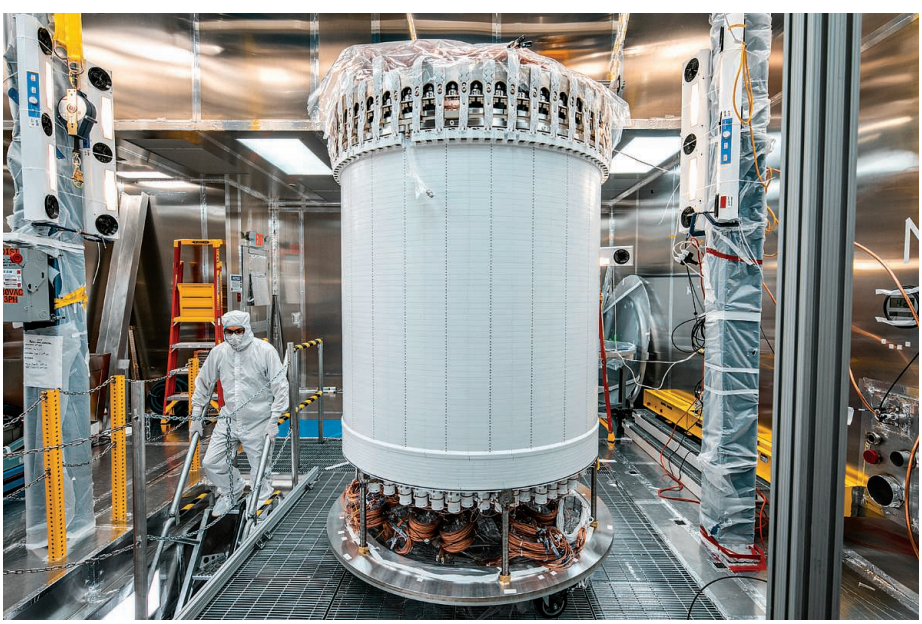
ASTROPHYSICS | The race to detect hypothetical particles of dark matter—the invisible stuff that binds together the galaxies with its gravity—enters a new phase this year with the startup of two powerful new underground detectors. Since the 1980s, physicists have used ever bigger and more sensitive ones to search for so-called weakly interacting massive particles (WIMPs), theorized to weigh 100 times as much as protons and to interact with other matter only through the feeble weak nuclear force. This year, the XENON-NT detector, which contains 8 tons of frigid liquid xenon, will turn on in the subterranean Gran Sasso National Laboratory in Italy. At the Sanford Underground Research Facility in South Dakota, the LUX-ZEPLIN (LZ) detector, which contains 10 tons of liquid xenon, will also power up. If XENON-NT and the LZ see nothing in the next few years, dark matter hunters could push for bigger WIMP detectors or set their sights on other hypothesized forms of dark matter. The Italian lab's future also remains uncertain, as former lab officials face prosecution for allegedly allowing contamination of local drinking water.

Making xenotransplants survive

BIOMEDICINE | The genome editor CRISPR is reinvigorating the beleaguered field of xenotransplantation, which aims to surgically replace human organs or tissues with ones harvested from animals such as pigs. Novel clinical trials of the strategy could launch this year. Xenotransplantation has long promised to alleviate a chronic shortage of human livers, hearts, and other organs. It could also provide corneas to cure blindness and insulin-producing islet cells to replace those destroyed by diabetes. But time and time again in earlier tests, human immune systems have quickly destroyed the foreign transplants. Recent CRISPR experiments have modified genes in pigs to prevent or dampen human immune responses to their tissue and have removed DNA from the porcine genome that could spawn potentially dangerous viruses in a person. Transplants from these edited pigs to monkeys, a key test of safety and efficacy before human trials, have demonstrated long-term viability in their new hosts.

Exascale computer to debut

COMPUTER SCIENCE | This year, China is expected to win the race to build the world's first exascale computer, capable of carrying out 1 billion billion (10^{18}) calculations per second, also known as an exaflop. Just which supercomputer will be the first remains uncertain, as China has set up a competition between three institutions: the National Supercomputing Center of Tianjin, the National Supercomputing Center in Jinan, and Dawning Information Industry Co., a manufacturer also known as Sugon. The new Chinese supercomputers, and others to follow in the European Union, Japan, and the United States, will be used to analyze vast data sets from astronomical and genetic surveys, and will support the continued rise of artificial intelligence. Some computer scientists expected the exascale milestone to have come sooner; delays resulted in part from the need to develop energy efficient computer chips.



The LUX-ZEPLIN dark matter detector is readied to record data at an underground lab in South Dakota.

PHOTO: MATTHEW KAPUST/SANFORD UNDERGROUND RESEARCH FACILITY

ALSO IN 2020

ALZHEIMER'S DRUG The U.S. Food and Drug Administration will decide whether to approve aducanumab, an antibody drug designed to bust the brain-clogging amyloid plaques of Alzheimer's disease. The experimental treatment has shown mixed success in clinical trials.

OCEAN CONSERVATION The United Nations intends to finish plans for a Decade of Ocean Science to begin in 2021. The goal is to coordinate work by scientists around the world to help improve ocean health. One expected emphasis is mapping more of the world's vulnerable marine ecosystems and biodiversity hot spots and more of the ocean's bottom, only about 4% of which has been charted in high resolution.

STEM CELL FUNDING California voters will decide in November whether to allocate $5.5 billion from bond sales to keep alive the California Institute for Regenerative Medicine. The funding agency was created through a $3 billion ballot initiative in 2004 to translate stem cell research into new therapies.

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

Published by AAAS

Aurorae were seen widely after Starfish Prime, a 1962 nuclear test in space.

SPACE PHYSICS

U.S. military tests radiation belt cleanup in space

Radio waves could sweep belts clean of satellite-killing particles after nuclear sneak attack

By **Richard Stone**

The U.S. military thought it had cleared the decks when, on 9 July 1962, it heaved a 1.4-megaton nuclear bomb some 400 kilometers into space: Orbiting satellites were safely out of range of the blast. But in the months that followed the test, called Starfish Prime, satellites began to wink out one by one, including the world's first communications satellite, Telstar. There was an unexpected aftereffect: High-energy electrons, shed by radioactive debris and trapped by Earth's magnetic field, were fritzing out the satellites' electronics and solar panels.

Starfish Prime and similar Soviet tests might be dismissed as Cold War misadventures, never to be repeated. After all, what nuclear power would want to pollute space with particles that could take out its own satellites, critical for communication, navigation, and surveillance? But military planners fear North Korea might be an exception: It has nuclear weapons but not a single functioning satellite among the thousands now in orbit. They quietly refer to a surprise orbital blast as a potential "Pearl Harbor of space."

And so, without fanfare, defense scientists are trying to devise a cure. Three space experiments—one now in orbit and two being readied for launch in 2021—aim to gather data on how to drain high-energy electrons trapped by Earth's magnetic field in radiation belts encircling the planet. The process, called radiation belt remediation (RBR), already happens naturally, when radio waves from deep space or from Earth—our own radio chatter, for example, or emissions from lightning—knock electrons trapped in Earth's Van Allen radiation belts into the upper atmosphere, where they quickly shed energy, often triggering aurorae.

"Natural precipitation happens all the time," says Craig Rodger, a space physicist at the University of Otago. But it would not nearly be fast enough to drain nuclear-charged radiation belts, where electron fluxes can be millions of times higher than in Earth's Van Allen belts.

Scientists got a glimpse of a potential solution from NASA's Van Allen Probes, which launched in 2012 and ducked in and out of Earth's radiation belts until the mission ended last summer. It offered a deep dive into natural remediation processes, showing how radio waves resonate with high-energy electrons, scattering them down the magnetic field lines and sweeping them out of the belts. "Compared to 10 years ago, we just know so much more about how these wave-particle interactions work," says Geoff Reeves, a space physicist at Los Alamos National Laboratory.

Now, researchers are ready to try artificial remediation, by beaming radio waves into the belts. Physicists have tested using the U.S. Navy's very low frequency (VLF) antenna towers, powerful facilities used to communicate with submarines, says Dan Baker, director of the Laboratory for Atmospheric and Space Physics at the University of Colorado, Boulder, and a lead investigator on the Van Allen Probes. The antennae of the High-frequency Active Auroral Research Program in Alaska and the giant dish of the Arecibo Observatory in Puerto Rico might also be enlisted to generate cleansing radio beams.

An orbiting RBR platform, closer to the target, could be more effective. In June 2019, the U.S. Air Force launched what it bills as the largest uncrewed structure ever flown in space: the DSX dipole antenna. Nearly as long as a U.S. football field, DSX's primary mission is to transmit VLF waves into the Van Allen belts and measure precipitating particles with onboard detectors. "It's a new way to prod the belts and explore basic questions in space physics," says DSX's principal investigator, James McCollough at

PHOTO: SCIENCE HISTORY IMAGES/ALAMY STOCK PHOTO

Published by AAAS

the Air Force Research Laboratory.

A team of scientists at Los Alamos and NASA's Goddard Space Flight Center is spearheading a second experiment in VLF precipitation. In April 2021, the team plans to launch a sounding rocket carrying the Beam Plasma Interactions Experiment, a miniature accelerator that would create a beam of electrons, which in turn would generate VLF waves capable of sweeping up particles. Reeves, who leads the experiment, believes the compact electron accelerator could ultimately be a better broom than a gigantic VLF antenna. "If we validate it with this experiment, we have a lot more confidence we can scale it up to higher power," he says.

A third experiment would coax the atmosphere itself to kick up turbulent waves that would draw down electrons. In the summer of 2021, the Naval Research Laboratory plans to launch a mission called the Space Measurements of a Rocket-Released Turbulence. A sounding rocket will fly into the ionosphere—an atmospheric layer hundreds of kilometers up that's awash in ions and electrons—and eject 1.5 kilograms of barium atoms. Ionized by sunlight, the barium would create a ring of moving plasma that emits radio waves: essentially a space version of a magnetron, the gadget used in microwave ovens.

The missions should help show which RBR system is most feasible, although an operational system may be years off. Whatever the technology, it could bring risks. A full-scale space cleanup might dump as much energy into the upper atmosphere as the geomagnetic storms caused by the Sun's occasional eruptions. Like them, it could disrupt airplane navigation and communication. And it would spawn heaps of nitrogen oxides and hydrogen oxides, which could eat away at the stratospheric ozone layer. "We don't know how great the effect would be," says Allison Jaynes, a space physicist at the University of Iowa.

Besides safeguarding against a nuclear burst, RBR technology could have a civilian dividend, Jaynes notes. NASA and other space agencies have long wrestled with shielding astronauts from the Van Allen belts and other sources of radiation on their way to and from deep space. VLF transmitters might be used to clear out high-energy electrons just before a spacecraft enters a danger zone. "When we become more active space travelers," she says, "it could provide a safe passage through the radiation belts." ■

Richard Stone is senior science editor at the Howard Hughes Medical Institute's Tangled Bank Studios in Chevy Chase, Maryland.

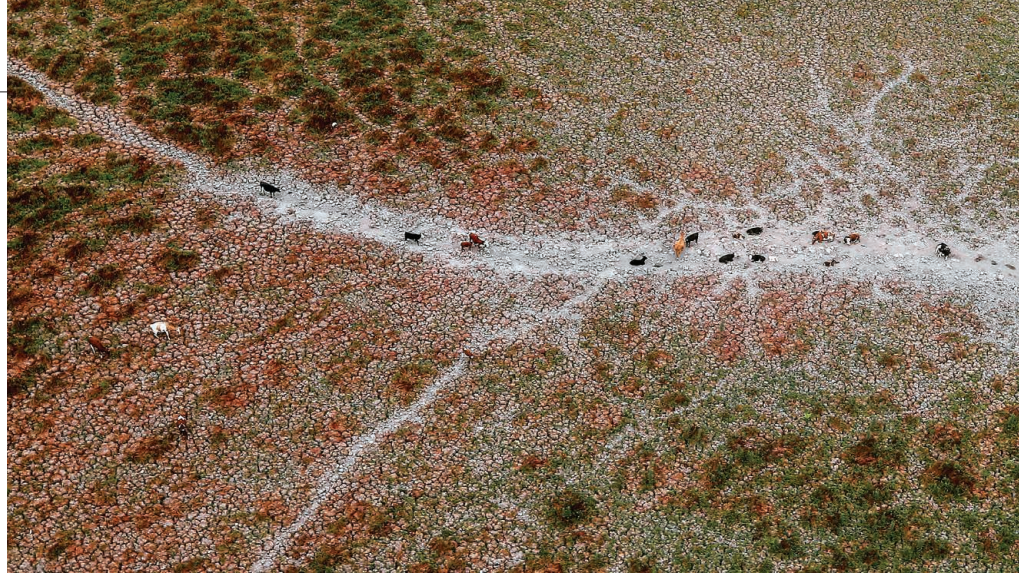


PHOTO: RODRIGO GARRIDO/REUTERS/NEWSCOM

CLIMATE

Past megadroughts hit North and South America in tandem

Strong La Niña conditions drove deep medieval droughts

By **Paul Voosen**

For 10 years, central Chile has been gripped by unrelenting drought. With 30% less rainfall than normal, verdant landscapes have withered, reservoirs are low, and more than 100,000 farm animals have died. The dry spell has lasted so long that researchers are calling it a "megadrought," rivaling dry stretches centuries ago. It's not so different from the decade-long drought that California, some 8000 kilometers away, endured until last year.

By analyzing tree ring records, scientists have now found evidence that such tandem droughts are more than a coincidence: They are surprisingly common over the past 1200 years, and they may often share a common cause—an abnormally cool state of the eastern Pacific Ocean known as La Niña. "We did not expect there to be as much coherence as we see," says Nathan Steiger, a paleoclimatologist at Columbia University who presented the work last month at a meeting of the American Geophysical Union. "They just happen together." The results suggest that, in the future, extreme aridity could strike all along the Americas' western coast.

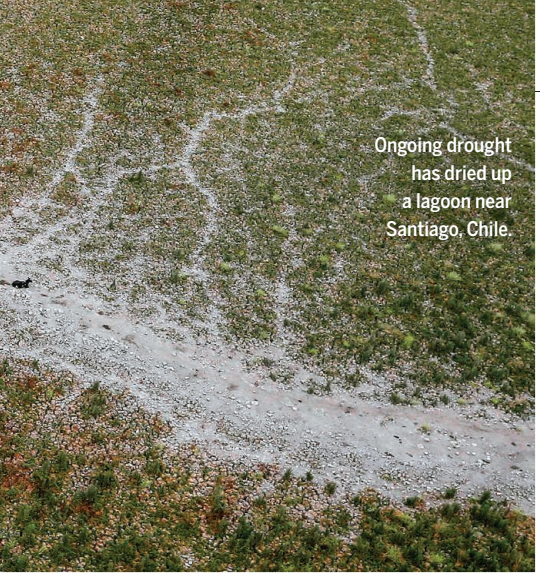
Evidence for synchronous, hemisphere-wide droughts first emerged in a 1994 study in *Nature*, which documented dead tree stumps in the middle of lakes and rivers in both Patagonia and California's Sierra Nevada. For trees to grow in stream- and lakebeds, the droughts must have lasted for decades, and at least one of these megadroughts seemed to have hit both continents simultaneously.

But pinning down the exact timing wasn't possible in the 1990s. Ancient drought is typically detected in variations in the width of tree rings, but tree ring records then were spotty. Since, however, tree ring scientists have put together "drought atlases" that provide consistent records for much of the world. "We'll have most hemispheric land areas covered by the end of the year," says Ed Cook, a tree ring scientist at Columbia.

Steiger combined these records with thousands of other proxies for dryness and temperature from trees, corals, ocean sediments, and ice cores, and fed them into a global climate model. Aligning itself to the records, it generated a global view of the changing climate, even in places with sparse proxies. The model confirmed that, from 800 to 1600 C.E., multiple megadroughts occurred simultaneously across the hemispheres. "It's there," Cook says. "Without question, it's there."

Besides correlating the varied climate records, the model also identified the key factors driving the climate variations. Steiger and his co-authors, including Cook, first used the new tool to look at megadroughts in the U.S. Southwest. Their study, published last year in *Science Advances*, is "amazing," says David Stahle, a tree ring scientist at the University of Arkansas in Fayetteville. "It's a bit like … they took those black-and-white films and colorized them." They found that megadroughts in the Southwest were influenced by three factors: an anomalously warm North Atlantic Ocean, small global temperature rises driven by factors such as

Published by AAAS



Ongoing drought has dried up a lagoon near Santiago, Chile.

a brightening Sun, and, especially, La Niña. The cold cousin of El Niño, La Niña can persist for years, deflecting rainstorms away from their usual tracks.

In the new work, the team finds that La Niña is almost the sole driver of the South American megadroughts. And because La Niña affects conditions on both sides of the equator, it could plausibly trigger simultaneous droughts in both hemispheres.

Jessica Tierney, a paleoclimatologist at the University of Arizona, says this marriage of proxy records and models is a powerful tool for understanding past climates. "Nathan has really been leading the way on that," she says. But she cautions that models don't perfectly simulate the La Niña cycle. It's also uncertain whether the link between La Niña and distant droughts is a stable dynamic that lasts centuries or might change over time. And the randomness of weather is always a factor: The current South American drought, for example, has endured through both El Niño and La Niña conditions.

Nor is it clear how the drought patterns will change as climate warms. A warming atmosphere alone seems certain to make megadroughts more frequent, especially in the Southwest. But scientists remain divided on how climate change will affect the El Niño-La Niña cycle. Models suggest El Niño will dominate, but in the past few decades, La Niña has seemed to be more frequent. "There are still camps," says Daniel Griffin, a paleoclimatologist at the University of Minnesota. "I see people trying to hold their noses for consensus statements."

As bad as the drought in Chile is today, it barely qualifies as a megadrought when compared with the medieval ones, which were longer and more severe. Clearly, there was something about that period that switched off in recent centuries, Cook says. If that pattern somehow came back, with greenhouse warming amplifying it, Cook says, "then things could get quite catastrophic." ■

PHOTO: KAI KUPFERSCHMIDT/*SCIENCE*

INFECTIOUS DISEASES

Study pushes emergence of measles back to antiquity

The virus may have entered the human population when cities grew large enough to sustain outbreaks

By **Kai Kupferschmidt**, *in Berlin*

On 3 June 1912, a 2-year-old girl at the Charité University Hospital here died of pneumonia following a measles infection. The next day, doctors took out her lungs, fixed them in formalin, and added them to a collection of anatomical specimens started by Rudolf Virchow, the "father of pathology." There they languished for more than 100 years—until Sebastien Calvignac-Spencer, an evolutionary biologist at the Robert Koch Institute, came across them in the basement of Berlin's Museum of Medical History.

Calvignac-Spencer and his team took a sample from the lungs, isolated RNA from it, and subsequently pieced together what is the oldest known genome of the measles virus. Its sequence helped them shed light on a much earlier period in measles' history. In a study slated for submission to the preprint submitted bioRxiv this week, the team concludes that the virus may have entered the human population as early as the fourth century B.C.E., rather than in medieval times, as previous research had suggested.

Scientists assembled a measles virus genome from the preserved lungs of a girl who died in 1912.

The work is technically brilliant, says evolutionary biologist Mike Worobey of the University of Arizona: "Just being able to get the measles virus out of these old, wet specimens. That sets the stage for all sorts of exciting work." Monica Green, a historian of infectious diseases at Arizona State University, Tempe, calls the sequencing "very impressive" as well but says the study lacks enough data points to "provide decisive answers" about measles' emergence. The authors agree. They hope sequences from antiquity, preserved in naturally mummified or frozen bodies, may one day do so.

Measles, which killed an estimated 142,000 people in 2017, is one of the most infectious human diseases. But when, where, and how it became a human pathogen is still debated. The closest relative of the measles virus is one that causes rinderpest, a disease that affected cattle, deer, buffalo, and other even-toed ungulate species before it was eradicated in 2011. Most researchers believe both viruses had a common ancestor that infected cattle. "The challenge is that … measles has left so few clear traces in historical disease descriptions," Green says.

Because measles spreads so fast and infection confers lifelong immunity, scientists estimate it needs populations of 250,000 to a half-million people to avoid burning itself out. Historians believe that the largest cities reached that size around the fourth century B.C.E. But when researchers in Japan used available genomes of the measles and rinderpest viruses to build a phylogenetic tree, enabling them to date the branches, they concluded in 2010 that measles didn't emerge until the 11th or 12th century C.E.

The uncertainty stems in part from a surprising lack of historic sequences. Only three genomes from measles viruses occurring before 1990 are known; the oldest is one isolated in 1954 that was turned into the first measles vaccine. So Calvignac-Spencer turned to the Berlin museum, whose shelves are filled with thousands of tissues and organs floating in formalin-filled glass cases, like aquariums of human anatomy.

Formalin fixes tissue by cross-linking

Published by AAAS

proteins and other large molecules, including RNA, which the measles genome is made of. To extract RNA from such samples, scientists use techniques pioneered about 10 years ago by cancer researchers interested in formalin-fixed biopsies. "We put them at 98° for 15 minutes and that breaks the cross-links," Calvignac-Spencer says. This also breaks up RNA, but modern methods allow scientists to sequence the fragments and piece them back together.

Calvignac-Spencer's team drew up a new phylogenetic tree using the 1912 genome as well as a new one from 1960, pieced together from a sample in another collection, and other available genomes. The resulting tree suggests the disease could have jumped to humans as early as 345 B.C.E.—right around the time human populations reached the critical size.

The earlier date for measles' emergence also reflects the models that the team used to analyze the viral sequences. When drawing up a family tree using differences in genomes, researchers must estimate the speed at which viral genomes diverge. In the past, their estimates were often too high, because some deleterious mutations tend to disappear over time. The new model accounts for this effect, called purifying selection. It pushes back the divergence of measles and rinderpest even without including the 1912 genome. But the genome strengthens the new timeline, Calvignac-Spencer says.

The researchers can't rule out that the measles virus first circulated in humans and then jumped to cattle, but that seems unlikely, says Albert Osterhaus of the University of Veterinary Medicine in Hanover, Germany. For one, ungulate herds probably reached the critical population size long before humans did. And the closest relative of the two viruses, which is even older, causes peste des petits ruminants, a sheep and goat disease that probably crossed to cattle more easily than to humans.

Similar studies have suggested that HIV and other pathogens also took off in the wake of major changes in the human population structure, Worobey says. "It seems like changes in human ecology really did coincide with the successful emergence of these viruses."

The ability to fish viral RNA out of very old samples has renewed interest in the Virchow collection, says Thomas Schnalke, head of the museum. "It came as a kind of revolution for us that researchers are coming and saying: 'Your samples are interesting for us again.'" Calvignac-Spencer has already labeled additional specimens he'd like to study with orange stickers. "It's a treasure trove," he says. "A window to the past that we can open now." ■

PHOTO: ARIZONA STATE UNIVERSITY

TRUMP ADMINISTRATION

Computer scientist in line to become next NSF director

Sethuraman Panchanathan would succeed France Córdova

By **Jeffrey Mervis**

The computer scientist whom President Donald Trump picked last month as the next director of the National Science Foundation (NSF) has followed the path taken by an untold number of foreign-born researchers by seeking greater opportunities in the United States. If the Senate confirms him, as seems likely, 58-year-old, India-born Sethuraman Panchanathan will become not only the second NSF director of Asian American descent, but a living embodiment of how the international flow of talent has helped fuel U.S. leadership in global science.

Panchanathan, who goes by "Panch," is executive vice president for research and chief innovation officer at Arizona State University (ASU) where he has worked since 1997. A former chair of its computer science and engineering department, he founded the university's school of computing and informatics and created and leads its center for cognitive ubiquitous computing, a technology that uses artificial intelligence and machine learning to help those with disabilities.

"His star has kept rising," says Subbarao Kambhampati, an ASU computer scientist and a longtime colleague and friend. "I think he's got three or four jobs. But he has a crazy amount of energy, and he seems to like going full speed all the time."

One of those jobs is being a member of the National Science Board (NSB), NSF's presidentially appointed oversight body. His 5-year stint has given him an insider's look at the $8 billion agency, and a chance to work with Trump's science adviser, Kelvin Droegemeier, who served on NSB together with Panchanathan for 2 years.

Droegemeier, who hailed the president's 19 December 2019 announcement of the pending nomination as "a win for science in the Trump administration," has strengthened ties between NSB and the White House Office of Science and Technology Policy since becoming OSTP director in January 2019. Some observers credit Droegemeier with not only championing Panchanathan for the NSF post, but also persuading the White House to queue up a new director before France Córdova finishes her 6-year term in March.

Panchanathan's interests dovetail with NSF priorities, says Kambhampati, who attended the same prestigious school, the Indian Institute of Technology (IIT), Madras, that launched Panchanathan's career. "He's very strong on interdisciplinary and use-inspired basic research, as well as entrepreneurship and partnerships with industry."

Curiously, Panchanathan will be the second IIT Madras alum to lead NSF; fellow alum Subra Suresh, a materials scientist, directed the agency for 2.5 years during the first term of former President Barack Obama. Suresh, who is 5 years older than

Published by AAAS

Sethuraman Panchanathan's research has focused on building technologies to help those with disabilities.

Panchanathan, attended IIT Madras in the mid-1970s, and he says its graduates—who are drawn from the most scientifically precocious of the country's vast pool of young people—"have done very well" in life.

Open borders fostered his own career, says Suresh, who left NSF in 2013 to become president of Carnegie Mellon University, and is now president of Nanyang Technological University. Leading NSF could put Panchanathan on a similar leadership trajectory. In 2017, he was runner-up to become president of the University of Arizona, and Kambhampati says he has been considered for other top academic positions. "Panch enjoys playing a leadership role within the scientific community," Kambhampati says. "He tends to have a big vision on issues."

Panchanathan declined to comment on his nomination. But in October 2019, he got the chance to appear on a national policy stage, testifying at a Senate hearing on "ensuring America's economic and strategic leadership" through research and innovation.

Senators spared Panchanathan when they pressed a panel of top scientists on the sensitive issue of combating threats to U.S. research from foreign influences. But he weighed in when they asked about ways to eliminate gender discrimination and remove barriers to entry for those groups underrepresented in science.

He cited a program at ASU that has attracted more undergraduate women into the mostly male discipline of engineering by showing how it relates to the social sciences, where women are often a majority. The social sciences also play an important role in his research. But he refused to offer legislators a silver bullet for creating a more diverse scientific workforce. Underrepresented groups "need more role models," he said. "And culture change is hard."

Even those who don't know him well expect him to be a good fit for NSF. "He's got all the right qualifications," says Arden Bement, an emeritus professor at Purdue University who was nominated by former President George W. Bush and preceded Suresh as NSF director. "And I think he'll be a good partner with Kelvin. … There's going to be a lot of uncertainty in the next year or so," Bement adds, drily, "and having those two hands at the helm should help science to stay on course."

Córdova, nominated by Obama to head NSF, says she's "thrilled with Panch's nomination." She deflected a question about her plans after leaving the agency, saying, "The most important thing for me is to finish strong." ■

CREDITS: (GRAPHIC) D. MALAKOFF/*SCIENCE*; (DATA) U.S. CONGRESS/WHITE HOUSE OFFICE OF MANAGEMENT AND BUDGET

U.S. BUDGET

Congress again rejects Trump cuts, smiles on science agencies

2020 spending bill increases research budgets, but next year will be tougher as lawmakers will have less to spend

By **Jeffrey Mervis**

Last summer, U.S. congressional leaders and the White House agreed to raise federal spending over the next 2 years by $125 billion above projected levels. The deal effectively removed the threat of large cuts to research agencies that President Donald Trump had proposed in his 2020 budget request. But those agencies didn't learn precisely how much money they would get until late last month, when Trump signed a $1.37 trillion spending bill for the 2020 fiscal year, which ends on 30 September.

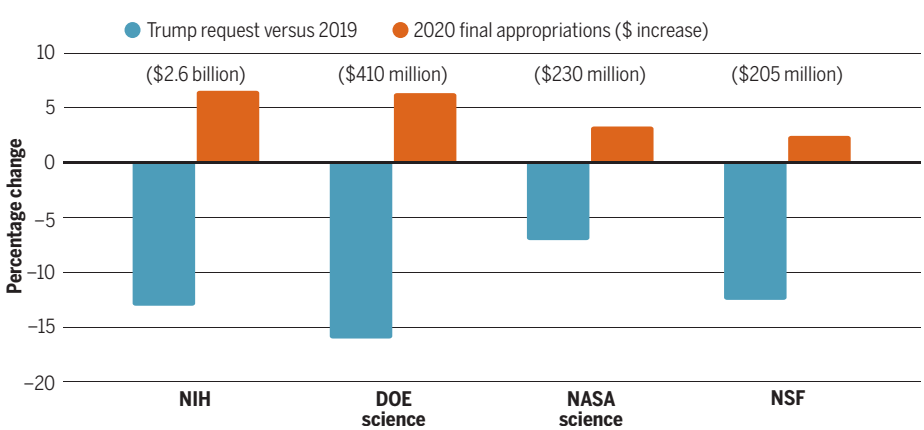
In the short term, the news is good. Once again, the National Institutes of Health (NIH) did the best, with Congress overriding Trump's proposed $5 billion cut and handing the agency a fifth straight annual increase of at least $2 billion. Science programs at the Department of Energy and NASA made a similar rebound from a proposed double-digit cut to a healthy increase, and DOE's Advanced Research Projects Agency-Energy escaped a proposed termination and instead will grow by a robust 16%. Congress also strengthened climate and environmental research programs at the National Oceanic and Atmospheric Administration and the U.S. Geological Survey that the Trump administration had targeted. And the bill includes the first federal funding for gun violence research in decades and language urging reductions in the use of animals in research.

However, this year's totals for research may be a high-water mark for Trump's 4-year term. Unless the White House and Democrats take the unlikely step of reopening the budget deal, Congress will have just $5 billion to add to agency budgets in 2021, far less than the $44 billion extra it was able to hand out this year. (The July 2019 agreement erased draconian cuts scheduled to go into effect in 2021.) So any major spending increases next year will have to come at the expense of cuts elsewhere.

The bad news came early for the National Science Foundation (NSF), which received a second straight year of modest growth. The agency was already under siege from Trump, who wanted to cut its $8 billion budget by 12%. Legislators had fought back: A spending bill passed by the House of Representatives in June 2019 would have grown NSF's budget by $565 million, and in October the Senate voted for a bump of $242 million. But rather than splitting the difference, Congress ended up giving NSF only a $205 million increase, less than both marks and only 2.5% above last year.

Reversal of fortune

Congress not only rejected deep budget cuts to major science agencies proposed by President Donald Trump for 2020, but also provided sometimes hefty increases in a final spending deal.



Published by AAAS

Competition with other programs funded within the same pot of money, together with a cut to that pot that came late in the process, is a major reason. Funding for the Executive Branch is spread across 12 appropriations subcommittees. The Commerce, Justice, and Science (CJS) panel funds NSF, as well as NASA and the Commerce and Justice departments.

Senator Richard Shelby (R–AL), chairman of both the Senate's CJS panel and the full committee, had already committed to giving the Commerce Department's Census Bureau the additional $3.5 billion it needed to conduct the 2020 census. Shelby is also a big supporter of NASA, which spends billions of dollars in his home state of Alabama. That huge economic impact translates into a vocal constituency.

In contrast, although NSF enjoys broad support among legislators, the agency "is everyone's second choice" when it comes to spending priorities, says Joel Widder, whose Federal Science Partners lobbies for universities and research institutions. So, after a 6-month negotiation between House and Senate negotiators left the CJS panels with less money than either had planned for, NSF lost out.

Research advocates say it's not such a bad deal. "An increase for NSF that is above inflation is a good thing and we should be thankful for it given all the other CJS pressures," says Joel Parriott of the American Astronomical Society. "And compared to the president's request, this is crazy good."

> **"An increase for NSF that is above inflation is a good thing and we should be thankful for it ..."**
>
> **Joel Parriot**, American Astronomical Society

One big winner within NSF's new budget are institutions planning new or upgraded "midscale research infrastructure," such as small telescopes or communications networks, that cost from $6 million to $70 million. Scientists say NSF funding has failed to keep up with the demand for these types of tools.

For the first time in decades, NSF this year requested $30 million for midscale projects in its research account, as well as $45 million in an account normally used to build more costly facilities. Congress liked the idea so much that it upped the amount in the large facilities account to $65 million, bringing the total for midscale projects to $95 million.

Lawmakers used the appropriations process to settle some policy matters. For instance, the final bill renews, for 10 years, a government-created nonprofit organization called the Patient-Centered Outcomes Research Institute (PCORI), which spends about $480 million per year on research comparing the benefits of medical treatments. Congress created PCORI under the 2010 Affordable Care Act, and patient advocacy and research organizations successfully pushed for PCORI's renewal (*Science*, 6 December 2019, p. 1179). Congress also gave the institute more leeway to consider costs in assessing competing treatments.

The bill also marks a turning point in the long-running debate over federal funding for research on gun violence. NIH and the Centers for Disease Control and Prevention will split $25 million, ending a 24-year drought for awards dedicated to such research. Congress also ordered the directors of NIH and CDC to report to Congress within 30 days on how it plans to support "ideologically and politically unbiased research projects." Scientists "spent so many years wondering whether we could even do this research. This is a clear signal that not only can we do it, but they want us to do it," says Charles Branas, a firearm violence epidemiologist at Columbia University.

Other provisions go beyond spending. At the urging of animal rights advocates, Congress ordered the Department of Veterans Affairs to devise a plan to reduce or end dog, cat, and primate testing by 2025. Similarly, it directed the Food and Drug Administration to develop a strategy and timeline for phasing out tests that involve primates and retiring its research monkeys, and NIH to tell Congress how it is moving to reduce the use of primates in research.

The budget also appears to settle, for now, a debate over whether the United States should remain fully engaged in ITER, the giant international experimental fusion reactor under construction near Cadarache in France. Some legislators have wanted DOE to withdraw from the project because of cost overruns and delays, and Congress had scaled back U.S. contributions. But lawmakers approved an 83% increase, to $242 million in 2020, including $85 million in cash and $157 million for manufacturing parts. That funding essentially puts the U.S. contribution back on the track DOE envisioned in 2017.

The new year means a new budget cycle. Trump will deliver his State of the Union address to Congress on 4 February and shortly after deliver his 2021 spending plan to Congress, which will then have its say. ■

With reporting by Adrian Cho, David Grimm, Jocelyn Kaiser, and Meredith Wadman.

PUBLIC HEALTH

Global polio eradication falters in the final stretch

Vaccine-derived outbreaks may force a change in "endgame" strategy

By **Leslie Roberts**

The "endgame" in the decadeslong campaign to eradicate polio suffered major setbacks in 2019. While the effort lost ground in Afghanistan and Pakistan, which recorded 116 cases of wild polio—four times the number in 2018—an especially alarming situation developed in Africa. In 12 countries, 196 children were paralyzed not by the wild virus, but by a strain derived from a live vaccine that has regained its virulence and ability to spread. Fighting these flare-ups will mean difficult decisions in the coming year.

The culprit in Africa is vaccine-derived polio virus type 2, and the fear is that it will jump continents and reseed outbreaks across the globe. A brand new vaccine is now being rushed through development to quash type 2 outbreaks. Mass production has already begun, even though the vaccine is still in clinical trials; it could be rolled out for emergency use as early as mid-2020. At the same time, the Global Polio Eradication Initiative (GPEI) is debating whether to combat the resurgent virus by re-enlisting a triple-whammy vaccine pulled from global use in 2016. That would be a controversial move, setting back the initiative several years, as well as a potential public relations disaster—an admission that the carefully crafted endgame strategy has failed.

"All options are on the table," says virologist Mark Pallansch of the U.S. Centers for Disease Control and Prevention, one of the five partner organizations in GPEI. "We are clearly in the most serious situation we have been in with the program," adds Roland Sutter, who recently stepped down as the director of polio research at the World Health Organization (WHO).

The heart of the problem is the live oral polio vaccine (OPV), the workhorse of the eradication program—the only polio vac-

Published by AAAS

Polio vaccinators in northwestern Nigeria, where the live-virus vaccine itself is causing new infections.

PHOTO: PIUS UTOMI EKPEI/AFP/GETTY IMAGES

cine powerful enough to stop viral circulation. Given as two drops into a child's mouth, OPV for decades contained a mix of three weakened polio viruses, one for each of the three wild serotypes that have long plagued humanity. All three serotypes in the vaccine have the potential to revert to more dangerous versions; that's why the endgame strategy calls for deploying OPV in massive campaigns to eradicate the wild virus, then ending its use entirely.

Wild serotype 2 was last sighted in 1999, so in 2016, as a first step in the endgame, all 155 countries using OPV replaced the trivalent version with a bivalent one, lacking the type 2 component. Announced with great fanfare, "the switch" was billed as the biggest vaccine rollout ever. Some type 2 outbreaks would inevitably occur for several years, GPEI realized, but those would be fought, somewhat paradoxically, by rushing in essentially the same vaccine that gave rise to them in the first place: a live, monovalent vaccine targeted against type 2 (mOPV2). If used in well-run campaigns, and only in outbreak regions, mOPV2 could stop outbreaks without seeding new ones, models suggested.

It often has not turned out that way. Instead of fading away, the number of type 2 outbreaks in Africa almost tripled from 2018 to 2019. Most of today's outbreaks stem from mOPV2 responses to previous ones, and GPEI is burning through its emergency stockpile of mOPV2 faster than it can be replenished. (Based on a small study in Mozambique, a WHO advisory panel recently recommended halving the dose to one drop if supplies run critically low, despite what it calls "a relatively weak level of evidence" that the smaller dose is as effective.) Meanwhile, the risk of explosive outbreaks around the globe is ratcheting up, because millions of children born since the switch have little or no immunity to type 2 virus.

WHO's Michel Zaffran, who leads GPEI, says there's room to make better use of mOPV2 by detecting outbreaks sooner, getting money and vaccines to countries earlier, and reaching more children. "There are things we can do even without a new tool," agrees Jay Wenger of the Bill & Melinda Gates Foundation, a partner in GPEI.

But hopes are pinned on a novel OPV (called nOPV2) that doesn't revert so easily. A Gates-funded research consortium is developing two candidates, each with changes at multiple nucleotides to increase genetic stability. Small phase I clinical trials suggested both trigger an immune response and are safe and unlikely to regain virulence. Phase II studies are underway in Belgium and Panama, but GPEI has already started to manufacture one candidate and hopes to have at least 100 million doses available this summer. GPEI is also pushing for an Emergency Use Listing, a never-before-used WHO mechanism that would enable the program to deploy the vaccine while it collects more data.

It's a risky strategy. The vaccine could fail or be delayed, and it won't solve all the problems. It won't be better at stopping outbreaks, just less likely to seed new ones. How much less likely remains to be seen. "Even if it is just 100 times safer, that will still be a big benefit," Wenger says, but the program is hoping for more.

Sutter worries GPEI is "putting all of its eggs into the nOPV basket." The novel vaccine could quickly lose its genetic stability if it exchanges key chunks of DNA with related viruses, he says. But how often these critical "recombination events" occur won't be known until the vaccine is used in larger populations. GPEI's Independent Monitoring Board noted recently that the program is "rather starry-eyed" about nOPV2's prospects.

If novel OPV2 doesn't work or vaccine-derived outbreaks spiral out of control before it is ready, the program might have little choice but to resurrect trivalent live vaccine, which would reintroduce immunity against type 2 in young children while maintaining protection against serotypes 1 and 3. The vaccine might be used in campaigns across Africa, reintroduced into routine immunization, or both.

The program is now struggling to define the "triggers" that would warrant this move. Is it reestablishment of type 2 across Africa? In Asia? The failure of nOPV2? The depletion of the mOPV2 emergency stockpile? "It is actually a hard question. … It's a public health judgment call," Wenger says. "People have different ideas on timing and triggers," Zaffran adds. But officials need to decide soon whether to ramp up production of OPV3 again, which could take several years.

Some experts fervently hope to avoid reintroduction of the trivalent vaccine. "It would be an enormous blow to the polio program and to international public health," says Nicholas Grassly, a modeler and epidemiologist at Imperial College London. Sutter, on the other hand, favors reintroduction sooner rather than later. Trivalent OPV "is the only thing we know has eradicated type 2 in the past and probably could eradicate it again," he says. But he agrees it would be a hard decision to communicate, given the huge global effort that went into persuading countries to switch to the bivalent vaccine in the first place. "How do we explain to the world that we have to go backward, not forward?" Sutter asks.

There's a bigger issue, too. No vaccine can stop polio if it doesn't get into children's mouths, program leaders and their advisers caution—and that has been a long-standing problem anywhere the virus, vaccine-derived or wild, still circulates. The polio eradication program has been struggling with complacency, fatigue, resistance, and poor planning—all human issues that technology can't fix. ■

Leslie Roberts is a journalist in Washington, D.C.

Published by AAAS

FIGHTING WORDS

Virologist Roberto Burioni has become a celebrity in Italy by sparring with vaccine skeptics

By **Douglas Starr**

In May 2016, Roberto Burioni, a virologist at Vita-Salute San Raffaele University in Milan, Italy, was asked to appear on a popular TV talk show to face off against two opponents of vaccines—a former DJ, Red Ronnie, and an actress and TV personality, Eleonora Brigliadori. The host gave most of the air time to the Italian celebrities. Then, with just a few minutes left, he turned to Burioni.

Burioni realized he didn't have time to make the usual arguments about statistics and scientific uncertainty, so he kept his message simple: "The Earth is round, gasoline is flammable, and vaccines are safe and effective," he said. "All the rest are dangerous lies."

"It went off like a bomb," Burioni recalls. Emails from viewers poured into the show, with some questioning how the publicly funded TV network hosting it could allow such ill-informed personalities to speak about medicine. Burioni took up the theme on his Facebook page, asking how one branch of government could broadcast lies about vaccines while its health agency promoted immunization. More than 5 million people responded to his comments. Radio journalist Alessandro Milan called Burioni's rebuttal to Red Ronnie "the 13 most beautiful words heard on TV in the last year."

So began the unlikely media career of Roberto Burioni. In just a few years, he has gone from being a respected but little-known professor to a major media personality and an internet savvy advocate for science. In a country where the government has sometimes promoted dubious medicine, such as unproven stem cell therapies, Burioni has become an outspoken advocate for scientific evidence on vaccines and other medical topics, and a harsh critic of pseudo-

PHOTO: TANIA/CONTRASTO/REDUX

Published by AAAS

Roberto Burioni had never appeared on TV before a 2016 appearance catapulted him to fame.

science. Nearly 480,000 people now follow him on Facebook—an impressive number in a country of 60 million. A web page he and colleagues established to provide general health information gets more than 100,000 visitors per month.

Burioni, with his shock of graying hair, peaked eyebrows, and ironic smile, appears often on TV and at public events. His four recent science books for popular audiences have become best sellers. The Italian edition of *Forbes* magazine named him one of Italy's top five internet game changers, and a former health minister nominated him for Italy's gold medal in public health.

Internet prominence brings trolls, and Burioni has been forced to worry about security. Some respected health researchers and journalists have also been critical, saying his blunt, even abrasive manner inflames an already polarized conflict. But many public health experts credit him with changing Italy's debate about vaccination and elevating the profile of science there.

"I think he's had a major impact on the public's understanding on the topic of vaccinations and science in general," says Pier Luigi Lopalco, who studies epidemiology and public health at the University of Pisa. "He's re-established the right of scientists to speak directly to the people without having a DJ or actor intervene."

BURIONI MIGHT seem an unlikely media personality. He followed a rigorous academic track: a medical degree in Rome; a Ph.D. in microbiology in Geneva, and several years of postdoctoral research at the U.S. Centers for Disease Control and Prevention, the University of Pennsylvania, the University of California, San Diego, and the Scripps Research Institute. As a professor in Milan, he develops monoclonal antibody therapies for herpes, hepatitis C, and other viral diseases—work that has led to 30 international patents.

He made his first foray into social media in 2015, when a friend who had created a Facebook group for mothers asked him to write an explainer about vaccines. Burioni, already irked by Italy's growing antivaccine movement, agreed. "I felt it was my duty to do something as a doctor and a professor and as a father of an 8-year-old daughter." He worried that the push to resist vaccinations could put her and her classmates at risk. He posted a five-point rebuttal of popular vaccine misconceptions and conspiracy theories—including the notion that drug companies promote vaccination in order to increase profits. "Pharmaceutical houses earn much more from disease cures than from vaccines," he declared. "So if you don't vaccinate your children, the pharmaceutical multinationals will be sincerely grateful to you."

That post and others hit a nerve. Within weeks his social media followers grew from about 100 to six digits. A voice like Burioni's—that of an expert who sees no reason to mince words or suffer fools—was evidently what many Italians were looking for.

As in many Western nations, concern about vaccines had surged in Italy in the late 1990s after U.K. doctor Andrew Wakefield published his now notorious study in *The Lancet* linking autism to the measles, mumps, and rubella (MMR) vaccine. The study was later shown to be fraudulent, and Wakefield lost his medical license. But that didn't stop him from continuing to proselytize against vaccines or prevent vaccine opponents from embracing his study.

> **"He's the one scientist who stood up and said, 'This is bullshit.'"**
>
> **Guido Silvestri**, Emory University

The Italian government insisted vaccines were safe, but a series of medical scandals had damaged its credibility. The most recent involved an experimental treatment called Stamina therapy, developed by an entrepreneur who claimed to be able to regenerate nerves from stem cells and cure conditions such as Parkinson's disease and muscular dystrophy. The inventor, Davide Vannoni, received huge fees from desperate patients, despite having produced no clinical studies or peer-reviewed papers. Scientist denounced the treatment, and the Italian Medicines Agency ruled it unsafe. But the Italian Senate, bowing to public opinion, permitted doctors and hospitals to administer it and funded a nearly $4 million clinical trial that was never completed.

Stamina therapy was finally outlawed in 2014, and Vannoni received a 22-month suspended sentence for fraud. (He died in December 2019.) Later, it was revealed that Italy's current prime minister, attorney Giuseppe Conte, won a court case in 2013 that allowed Stamina therapy to be given to a girl suffering from an incurable neurological condition. The child later died.

Courts in Italy have compounded the problem. Local judges often lack scientific expertise and are allowed pick their own consultants on technical matters. In one egregious case in 2012, a judge in Rimini relied on Wakefield's discredited research to award the family of an autistic boy more than $200,000 from the government, on the grounds that the measles vaccine distributed by the national health service caused his condition. News of the decision spread on the internet like a contagion, igniting false beliefs and conspiracy theories. "The year 2012 was identified as the breaking point in the public's confidence in vaccination in Italy," researchers at the University of Pisa wrote in a study of the web's effect on vaccinations. An appeals court reversed that judge's decision in 2015, but the damage was done.

Antivaccination sentiment infected not only the courts, but also entertainment and politics. In the late 1990s, for example, a comedian named Beppe Grillo had become famous in Italy for denouncing vaccination. Ten years ago, he co-founded the Five Star Movement, a libertarian political party that became a dominant member of Italy's ruling coalition and embraced antiscience positions.

Vaccination rates, which had been climbing since the 1990s, started to slide. Uptake of the MMR vaccine declined from a peak of nearly 94% in 2010 to just over 85% in 2015—one of the lowest rates in Europe, and well below the 95% needed for herd immunity. Almost in lockstep, the nation's measles rate climbed to the second highest in Europe, after Romania's.

THAT WAS THE SITUATION Burioni waded into when he made his appearance on Italian TV. He'd never been on TV before, and when the network called he assumed he'd be speaking to other medical experts. He was shocked to find that he'd be sharing the broadcast with two people who knew nothing about vaccines. Hence his curt reply, and the public's enthusiastic response.

As he saw it, a door had opened and he had to walk through it. The public seemed hungry for straight talk from an expert, and he obliged. "I realized that the language of social media needs to be different than the language used in conferences, with colleagues or even with patients," he says, "so I tried to use not a single difficult word." He turned instead to concrete, nonmedical metaphors in his Facebook and Twitter postings. "Does an aircraft engineer take a vote among the passengers as to how many wheels to put on an airplane?" he asked. "No—the engineer is the expert, he's trained for this job, and it's his job to decide."

When one of the largest publishers in Italy asked him to write a book on vaccines, Burioni cranked it out in 4 months. A night owl, he wrote after his wife and daughter went to bed. "I never realized that I was good at writing," he says. "Here I was at 54 years old; how would you say it—a late bloomer?"

Published by AAAS

A 2017 law requiring childhood vaccinations triggered protests and threats in Italy. In the Senate, Roberto Burioni's allies displayed a doctored photo that circulated online, showing his face grafted onto an image of Aldo Moro, an Italian prime minister murdered in 1978. In Rome, opponents of the law flooded the streets.

PHOTOS: (TOP TO BOTTOM) AUGUSTO CASASOLI/A3/CONTRASTO/REDUX; STEFANO MONTESI/CORBIS VIA GETTY IMAGES

The result, *Vaccines are not an Opinion: Vaccinations Explained to Those Who Really Don't Want to Understand*, was a novelty in a country where scientists rarely communicate in colloquial language. In it, Burioni explains how vaccines work, traces the history of vaccination and of vaccine denialism, and dismantles the denialists' arguments. He brings up tragic case histories of people who died young for want of a vaccine and skewers prominent vaccine opponents, such as Jenny McCarthy, who declared on *Oprah* that she learned her biology "from the university of Google." He hammers on the theme that you can't substitute opinion for facts. The book's popularity made him a fixture on radio and TV. "The speed of light can't be decided by a show of hands," he'd say in talk show appearances.

In his later bestsellers, Burioni widened his case against pseudoscience in Italy. "We are a country forever wobbling between science and superstition," he says. He railed against judges who ruled that unproven treatments such as Stamina therapy had to be given to patients who requested them, often in publicly funded hospitals. He also criticized the Italian public health system for reimbursing patients for homeopathy—pseudoscientific medical treatments that use an extremely dilute concentration of a substance that causes symptoms similar to those of the disease.

Burioni says he wants to promote a respect for expertise. "I know something about vaccines, viruses, and bacteria because I have been studying them for a lifetime," he writes in *Conspiracy of Dunces: Why Science Can't be Democratic*, his second book. "But I have no idea how to bake a cake or wire a lamp, so I go to a bakery or call an electrician."

Trouble begins, he says, when electricians, bakers, and other nonscientists feel qualified to weigh in on vaccination. The internet abets the problem, he says: Unfiltered by editors, it levels the playing field between experts and "dunces." A mass media committed to presenting both sides of every issue makes things worse. "I can't support a world in which lies are given the same dignity as the truth," he heatedly remarked on an interview show. "Enough already! Enough!" The audience burst into applause.

Burioni's books, postings, and media presence have made him a celebrity scientist in Italy akin to, say, Neil deGrasse Tyson in the United States. Hardly a week goes by without him appearing on TV or in a glamour shot in a newspaper or magazine. At a recent conference in Milan about future technologies, the audience—including many influential business leaders—swarmed him after his talk, asking for advice and autographs.

Published by AAAS

"I feel a bit embarrassed by it all," Burioni says, seeming not at all embarrassed but very much amused. (In fact he seems to relish the publicity, having recently appeared on an Italian comedy show in a version of *Name That Tune*.) He's proud that despite all the media appearances he hasn't missed giving a lecture and continues to work full-speed in his lab.

Yet it hasn't all been fun. His time for family and leisure has suffered. More disturbing, at one point police had to stand watch over his house after someone threatened his daughter online—one of many death threats. In 2018, Burioni and his family were vacationing at the beach town of Rimini when a vaccine opponent caught sight of him and posted on the web. Within minutes an online posse had weighed in with suggestions on how to harass him; on the advice of police he and his family went elsewhere.

He's also been the subject of more substantive criticism for his "Burioni blasts"—devastating replies to even the mildest opposition. When a Facebook follower said he was trying to figure out the vaccine issue in his own head, Burioni replied, "When you go on social media you can be reasonably sure that the interior of that head is as empty as a tire tube." He's also been known to ban even mild critics from his Facebook page and discount the work of science journalists on the vaccine issue. "This is not what you would call public outreach," says Sergio Pistoi, a science writer and molecular biologist in Tuscany.

Burioni's broadsides polarize and oversimplify a complicated discussion, adds Fabio Turone, a science writer and director of the Center for Ethics in Science and Journalism. Fewer than 1% of Italians are hard-core vaccine deniers, he says; but about 15% are "vaccine hesitant" because they have concerns about vaccines or find them hard to get. (In Italy, doctors at the public health service—not the child's pediatrician—administer vaccinations, and some health service centers closed after the budget crash of 2008.) Such people are best reached by persuasion, not mockery, Turone says.

Roberta Villa, a journalist and doctor, calls for a more empathetic approach. Her YouTube videos show her sitting at her kitchen table, a cup of coffee in hand and a drawing of Snoopy in the background, explaining that as a mother of six she understands that all parents want to protect their children. That's why scientists have examined vaccines with such care, she tells viewers. "People are much more educated than they used to be," she says. "They want to understand what we are doing to their children. So you cannot approach them in a paternalistic way."

Others appreciate Burioni's no-holds-barred style. "I know he likes to make jokes that sometimes can be seen as abrasive," says Guido Silvestri, a longtime friend and a pathologist at Emory University. "But what kind of debate can you have with someone who says that vaccines are a conspiracy to kill children? He's the one scientist who stood up and said, 'This is bullshit.'"

THINGS ARE IMPROVING. In the summer of 2017, after the nadir in vaccination rates and the spike in preventable disease, the government passed a law that set up public education programs and requires all schoolchildren to get 10 essential vaccinations before kindergarten. Even though the law is loosely enforced, it's working. Vaccination rates have rebounded, according to Italy's National Institute of Health. After bottoming out at 85.3% in 2015, measles vaccination has risen to 94.1%, within striking distance of the 95% needed to prevent outbreaks when single cases pop up. The law has proved so effective that France and Germany passed similar laws in 2019.

Burioni and others don't claim he is responsible for the turnaround, but public health researchers say the "Burioni effect" has altered public discourse. Four years ago, googling "vaccines" in Italy would bring up a list of antivaccination groups among top results. Now, the first hits include Burioni's websites and the vaccine information sites created by the World Health Organization and Italy's health authorities. Similarly, radio and TV shows have become more likely to book scientists instead of actors and DJs to discuss vaccination, according to Lopalco and others who monitor the media.

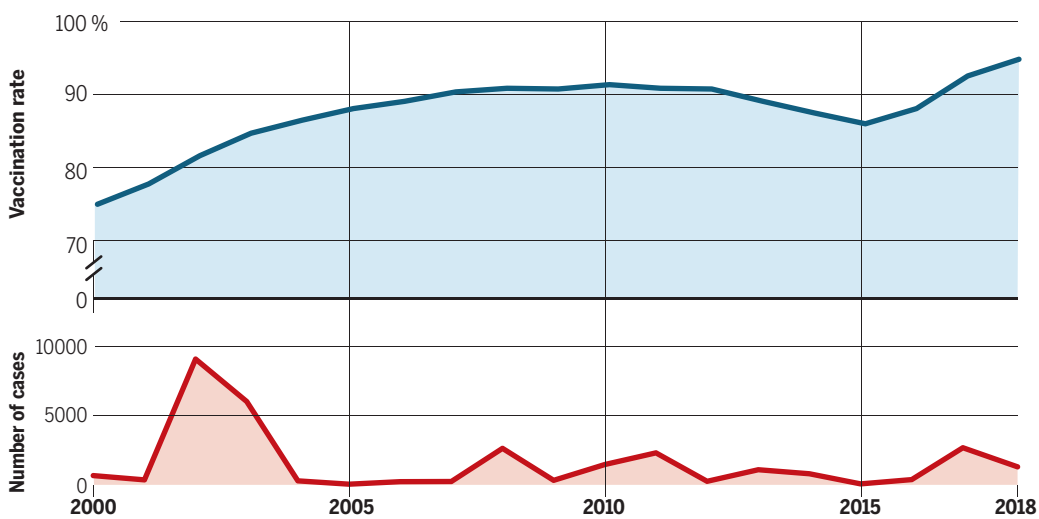
Recently, Burioni expanded his campaign for science. In late 2018, he and half a dozen colleagues created Medical Facts, a web portal that posts news, advice, and comments on a variety of health issues. In January 2019, he and Silvestri posted a manifesto called "The Pact for Science," supporting research and education and calling for the use of objective science in government decisions. It has attracted thousands of signatures from people as diverse as the head of the left-leaning democratic party and Grillo himself, one of Italy's original vaccine deniers. In June 2019, Burioni and colleagues created a group related to the pact that will advocate for science and help local judges gather better information when adjudicating cases involving science and health.

"The world has changed," Burioni says, acknowledging both the good and "catastrophic" effects of social media. "Science needs to find a new voice—not the language of scientific congresses, but a language that's understandable, passionate, and convincing." ■

Douglas Starr is a journalist in Boston.

Setback and recovery

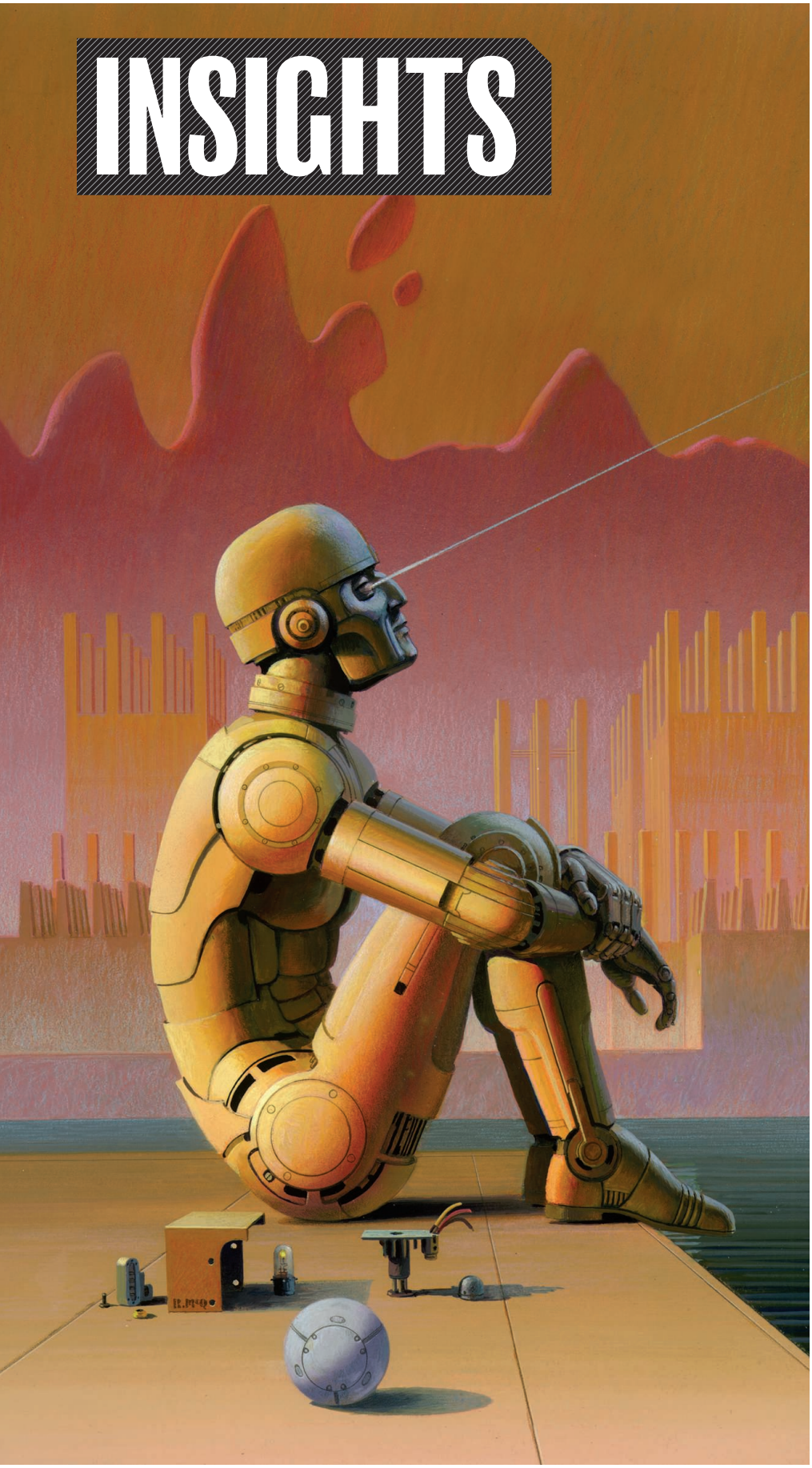
After rising for years, measles vaccination rates in Italy fell until 2015 because of unfounded safety concerns, abetted by some government and court actions; measles cases spiked. Public education and a 2017 law have boosted rates since then.



CREDITS: (GRAPHIC) N. DESAI/*SCIENCE*; (DATA) EPICENTRO; F. D'ANCONA *ET AL*., *EURO SURVEILL*. 24(26). (2019); *INTEGRATED SURVEILLANCE OF MEASLES AND RUBELLA IN ITALY*, 55, (2019)

Published by AAAS

INSIGHTS



BOOKS *et al.*

SCIENCE FICTION

Asimov at 100

From epic space operas to rules for robots, the prolific author's literary legacy endures

By **James Gunn**

Gertrude Blugerman once asked of her husband, Isaac Asimov, "What will you say at the end of your life if you have written one hundred books but have missed living?" "Only one hundred?" he replied. Asimov wrote of this incident in *Opus 100*, his hundredth book, published in 1969. He would write or edit more than 500 books in his lifetime.

Asimov was brought to the United States at the age of 3 from Petrovichi, a small town in the still relatively new Soviet Union. His birthday—2 January 1920—was one settled on by his parents. (Records in Petrovichi were unreliable, and he may have been born as early as November 1919.) The family settled in Brooklyn, New York, where his father opened a candy store that also sold newspapers and magazines. It was an enterprise in which all members of the family participated. Perhaps not surprisingly, Isaac was drawn to the magazines, particularly the colorful science fiction publications.

Asimov was a gifted student with a capacious memory who moved easily through grade school and high school. In the first of his autobiographies (he published two comprehensive volumes and, later, two supplemental tomes), he recounted how he would acquire his textbooks for the semester, read them on his walk home, and never open them again. When I interviewed him for the book I wrote about his life and work (*1*), I asked if he ever forgot anything, and he said that once he had been reciting the second verse of the U.S. national anthem and, for a moment, could not think of how it started.

Asimov attended Seth Low Junior College, a branch of Columbia University in Brooklyn, graduating in 1939 from Columbia when Seth Low closed in 1938. He majored in chemistry after discovering in his freshman year that to major in zoology, one would be required to dissect cats.

This pensive android appeared on the cover of Asimov's 1990 collection *Robot Visions*.

PHOTO: COURTESY OF THE RALPH MCQUARRIE ARCHIVES

Published by AAAS

The year 1939 was also when Asimov's first science fiction story was published in the magazine *Astounding Science Fiction*. The sale of his stories paid for his college expenses, including the master's degree he would later earn in chemistry. (He was rejected twice for medical school but would go on to earn his doctorate, again in chemistry.)

In 1942, Asimov began research as a chemist at the Philadelphia Navy Yard. Meanwhile, he continued to sell stories but considered himself a third-rate writer until his novelette "Nightfall" received a cover story in *Astounding Science Fiction* in 1941. The following year, the first story of his *Foundation* series was published. During this period, he began the first of his robot stories, which were published together in 1950 as *I, Robot*.

In September 1945, Asimov was drafted into the U.S. Army and served for 6 months before being honorably discharged. He returned to Columbia, where he earned his doctoral degree in 1948 before accepting his first (and only) academic appointment at Boston University.

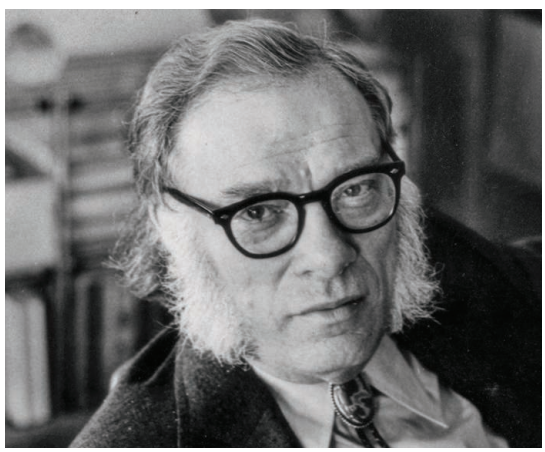
Science fiction received a boost from World War II. The number of magazines publishing it increased in the aftermath of the war, and book publishers soon followed. The subsequent Space Age and concerns evoked by Sputnik also led to publishing opportunities in the realm of science popularization. Asimov's engagement with the latter genre began with a 1953 biology text coauthored with two other faculty members. This would be the beginning of a series for which the earnings—together with those of his science fiction and other writing—soon exceeded his university salary.

When Asimov was reprimanded for failing to conduct any research, he replied that he considered his writing his research. Shortly afterward, he was terminated. He had previously been promoted to the rank of associate professor with tenure, and after a 2-year battle it was determined that he would be allowed to stay on. He chose instead to leave to focus on his writing.

By the time I filmed him for my literature of science fiction film series in 1972, Asimov was living in a high-rise off of Central Park and was in the middle of divorcing his first wife. "Science fiction writers and readers didn't put a man on the moon," he reflected during our interview, "but they created a climate of opinion in which the goal of putting a man on the moon became acceptable."

Asimov had, by this time, already published his hundredth book—the previously mentioned *Opus 100*—and was, it turned out, only just getting started. Soon his books, written on topics ranging from the Bible to the human body, began appearing as often as monthly, leading Harvard paleontologist George C. Simpson to call him "one of our natural wonders and national resources" (*1*).

Asimov was a popular public speaker and a regular participant in science fiction conventions, where, despite having earned a formidable reputation as one of the "big three" science fiction authors of the era, he remained approachable. He participated in a variety of social organizations, including Mensa; the Humanist Society (of which he was named the honorary president); the Baker Street Irregulars, a society dedicated to the appreciation of Sherlock Holmes; and the Trap-Door Spiders, a luncheon group that served as the inspiration for the fictional Black Widowers club in a series of mystery stories and novels he wrote.



Asimov brought drama and narrative to his nonfiction science writing and often grounded his fiction in real scientific principles.

Kurt Vonnegut is reported to have once asked Asimov how it felt to be the man who knows everything, to which Asimov is said to have replied that he only knew how it felt to have the reputation of omniscience. On another occasion, an editor reportedly encouraged Asimov to write an autobiography. "But I've never done anything," he protested. He returned a year later with a thick manuscript. When the editor failed to protest its length, Asimov left and returned with another manuscript, just as thick. "What would you have written if you had done anything?" the editor is said to have replied.

Although he claimed no false modesty—"nor true modesty either" (*2*)—he insisted on acknowledging the role of others in his success. He credited editor John W. Campbell with the invention of the three laws of robotics and with introducing him to the Ralph Waldo Emerson quotation that inspired Asimov to write "Nightfall": "If the stars should appear one night in a thousand years, how would men believe and adore; and preserve for many generations the remembrance of the city of God which had been shown! But every night come out these envoys of beauty, and light the universe with their admonishing smile." And in an article in *Isaac Asimov's Science Fiction Magazine* (*3*), Asimov credited my words with helping him overcome his fears about writing a sequel to the *Foundation* trilogy. The resulting series (*4*) brought the *Foundation* trilogy and his robot novels together into a single future history and became his first bestseller.

A case can be made that, like H. G. Wells, Asimov came along at the right time. (Wells once commented that he made his writing debut in the 1890s, when the public was looking for new writers.) But Asimov also had a restless and productive mind. His early experience of reading, and then writing, science fiction gave his popular science writing a rare narrative model, while his fiction similarly benefited from his scientific training.

Some of Asimov's critics complained that his writing lacked style. He responded by asserting that he had a style: clarity. But it also was true that he was able to adopt new methods, particularly in his later works.

Asimov's fiction was based on the presumption that humanity would solve its problems by thinking coolly and logically. In his nonfiction writing, he often grappled with the messier realities of human nature. There are no records of how many minds he influenced with the latter, but his ability to communicate difficult scientific ideas in simple language has not been equaled since.

Asimov once told a friend that if he had a hang-up, it was his desire to write, saying that he wanted to die with his nose stuck between two typewriter keys. It did not happen that way. His second wife and widow, Janet, confirmed after his death that Asimov had contracted AIDS from a blood transfusion during open-heart surgery a decade before. He had been persuaded by his doctors to keep this information confidential, because of concerns that it would deter people from undergoing necessary surgery. He died from complications of the disease in 1992 at the age of 72. But his legacy, and his books, remain. ■

The reviewer is the founder of the Gunn Center for the Study of Science Fiction, University of Kansas, Lawrence, KS 66045, USA. Email: jgunn@ku.edu

PHOTO: MONDADORI PORTFOLIO/CONTRIBUTOR/GETTY IMAGES

REFERENCES AND NOTES

1. J. E. Gunn, *Isaac Asimov: The Foundations of Science Fiction* (Oxford Univ. Press, 1982).
2. J. E. Gunn, *Alternate Worlds: The Illustrated History of Science Fiction* (A & W Visual Library, 1975).
3. I. Asimov, "The story behind the 'Foundation,'" *Isaac Asimov's Science Fiction Magazine*, December 1982.
4. I. Asimov, *Foundation's Edge* (Doubleday, 1982).

10.1126/science.aba0303

Published by AAAS

PERSPECTIVES

INFECTIOUS DISEASE

Uncovering the ART of antimalarial resistance

A key mechanism of resistance to the antimalarial drug artemisinin is identified

By **Danushka Marapana** *and* **Alan F. Cowman**

The identification of artemisinin (ART) in 1971 allowed treatment of malaria resistant to chloroquine, the prevailing drug at the time, and provided hope for a malaria-free world (*1*). Today, malaria control efforts have been very successful, with 32% fewer deaths over the past 8 years (*2*). However, the emergence of resistance to ART and other antimalarials threatens to become a major problem in the continuing program to eliminate and eventually eradicate malaria (*3*). ART combination therapies (ACTs) are the current gold standard for the treatment and control of malaria, and how parasites that cause malaria in humans mediate this resistance is of intense interest for preventing the spread of drug resistance. On page 51 of this issue, Birnbaum *et al.* (*4*) answer this critical question by identifying the molecular mechanism of ART resistance in the most lethal human malaria parasite, *Plasmodium falciparum*.

As malaria parasites infect and grow within human erythrocytes (red blood cells), they actively engulf and digest host hemoglobin in a dedicated parasite-derived food vacuole (see the figure). Hemoglobin is broken down into amino acids, which are used for the synthesis of parasite proteins, and iron-bound heme, which is gradually detoxified within hemozoin crystals. Heme binding results in cleavage of the endoperoxide bond of ART, enabling drug activation. Active ART produces free radicals and reactive oxygen species that attack protein and lipid molecules of the developing parasite. ART treatment of intraerythrocytic *Plasmodium* parasites results in rapid killing of all life cycle stages, including the young "ring" stage of infection, which is resistant to most antimalarial drugs (*5*). However, the initial identification of ART resistance in western Cambodia in 2008 followed by rapid spread throughout the Greater Mekong Subregion of Southeast Asia forewarned of a major issue for the malaria elimination agenda in that region (*3*). Additionally, the potential spread of ART resistance to sub-Saharan Africa would make elimination programs even more challenging.

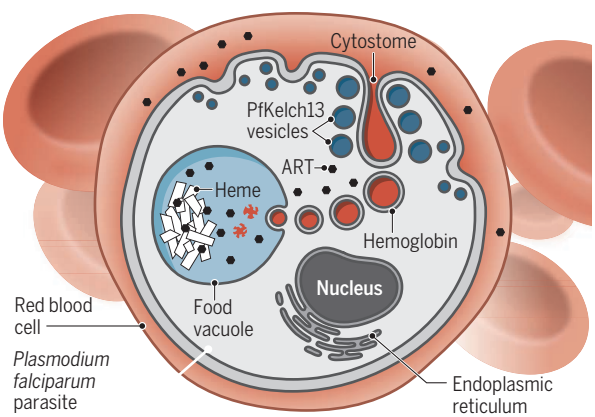
ART resistance manifests as enhanced survival and delayed clearance of young ring-stage parasites after a concentrated exposure to ART or its derivatives. Molecular genotyping of ART-resistant parasites uncovered single nonsynonymous mutations in a *P. falciparum*–specific gene called *Pfkelch13* (*6*). The PfKelch13 protein contains three main functional regions: a parasite-specific localization sequence, a BTB/POZ domain that typically facilitates ubiquitin-mediated degradation, and a carboxyl-terminal Kelch propeller repeat region that is predicted to function as a scaffold for protein-protein interactions (*6*). Nearly all clinically relevant ART-resistance mutations are localized within this carboxyl-terminal Kelch-repeat region. Directed mutagenesis studies of both ART-naïve strains and clinical isolates have confirmed the causal role of PfKelch13 in mediating ART resistance (*7*, *8*). Multiple hypotheses implicate PfKelch13 as a responder to downstream effects of ART activation, especially in up-regulation of pathways involved in the cellular stress response and reduced protein translation in the presence of ART-induced stress (*9*–*11*). However, the possibility of PfKelch13 acting upstream as a conduit for ART activation was not investigated prior to the Birnbaum *et al.* study.

Birnbaum *et al.* found that PfKelch13 and its interacting proteins are localized in vesicles close to cytostomes, which are erythrocyte-cytosol containing structures produced by the parasite. They discovered that proteins in this PfKelch13 compartment are required for the endocytic uptake of hemoglobin; however, PfKelch13 itself is essential for this process only in the young ring-stage parasites. By investigating the link between PfKelch13, hemoglobin uptake, and ART activation, Birnbaum *et al.* found that PfKelch13-inactivated parasites and those carrying ART resistance–conferring *Pfkelch13* mutations display depleted concentrations of the protein. These PfKelch13-depleted parasites exhibit decreased hemoglobin uptake as well as enhanced ring-stage survival during high-dose ART treatment. Thus, PfKelch13 is a facilitator of ART activation, and Birnbaum *et al.* propose an elegant model to merge a large body of prior research into a comprehensive pathway of ART resistance.

In this model, mutations in the Kelch propeller domains of PfKelch13 impair the ability of parasites to endocytose hemoglobin during the young ring stage of infection. Consequently, hemoglobin degradation is reduced, and less heme becomes available for ART activation. Hemoglobin catabolism also provides the parasite with a source of amino acids for protein synthesis. Therefore,

Gateway for activation

Malaria parasites ingest hemoglobin to produce heme from the host red blood cell using cytostomes. The interaction of heme in the parasite food vacuole with artemisinin (ART) causes activation of the drug and parasite killing. PfKelch13-containing vesicles regulate the uptake of hemoglobin by *Plasmodium falciparum* and thereby affect the amount of ART activation through reduced heme availability.



GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

Department of Medical Biology, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Parkville, VIC 3052, Australia. Email: cowman@wehi.edu.au

Published by AAAS

it seems that resistance to ART carries a biological trade-off for the parasite: It grows at a slower rate owing to the decreased availability of amino acids for protein translation. Therefore, the unfolded protein response, which alleviates cellular stress caused by accumulation of unfolded and misfolded proteins, is up-regulated in parasites in an attempt to withstand the toxic effects of reactive ART metabolites (*10*, *11*).

How does PfKelch13 function in hemoglobin endocytosis? A likely scenario is that the cysteine-containing Kelch domains, which allow protein-protein interactions, function as a molecular scaffold to engage PfKelch13 partner proteins to facilitate hemoglobin transport. Inhibitors against eukaryotic Kelch proteins disrupt protein-protein interactions by chemically modifying the Kelch cysteine residues (*12*). This could be an avenue to disrupt PfKelch13-mediated hemoglobin digestion and nutrient uptake by the parasite. A relatively unexplored facet of PfKelch13 is the BTB/POZ domain (*6*). In other eukaryotes, this domain forms a Cullin-3–E3 ubiquitin ligase complex and promotes the proteasomal degradation of multiple protein substrates (*13*). However, *P. falciparum* does not appear to encode a canonical homolog of Cullin-3, and because PfKelch13 itself is localized within a distinct subcompartment in the parasite cytoplasm (*14*), it will be interesting to identify how PfKelch13 ubiquitinates proteins.

Pfkelch13 mutation is not the only mechanism of ART resistance, because parasites that display an increased ring-stage survival phenotype but lack *Pfkelch13* mutations have been reported (*15*). It would be intriguing to investigate whether these parasites also show defects in hemoglobin uptake. In addition to ART, ACTs include a partner drug that typically targets the hemoglobin digestion mechanism of the parasite. Future formulations of ACTs might be more effective by using partner drugs targeting other biochemical pathways. ■

REFERENCES AND NOTES

1. Y. Tu, *Nat. Med.* **17**, 1217 (2011).
2. "World malaria report 2018" (World Health Organization, 2018); www.who.int/malaria/publications/world-malaria-report-2018/en/.
3. A. M. Dondorp *et al.*, *N. Engl. J. Med.* **361**, 455 (2009).
4. J. Birnbaum *et al.*, *Science* **367**, 51 (2020).
5. C. L. Hartwig *et al.*, *Biochem. Pharmacol.* **77**, 322 (2009).
6. F. Ariey *et al.*, *Nature* **505**, 50 (2014).
7. J. Straimer *et al.*, *Science* **347**, 428 (2015).
8. M. Ghorbal *et al.*, *Nat. Biotechnol.* **32**, 819 (2014).
9. C. Dogovski *et al.*, *PLOS Biol.* **13**, e1002132 (2015).
10. S. Mok *et al.*, *Science* **347**, 431 (2015).
11. P. J. Shaw *et al.*, *BMC Genomics* **16**, 830 (2015).
12. A. Cuadrado *et al.*, *Nat. Rev. Drug Discov.* **18**, 295 (2019).
13. L. Pintard, A. Willems, M. Peter, *EMBO J.* **23**, 1681 (2004).
14. J. Birnbaum *et al.*, *Nat. Methods* **14**, 450 (2017).
15. A. Mukherjee *et al.*, *Malar. J.* **16**, 195 (2017).

10.1126/science.aba0445

GRAPHIC: C. BICKEL/*SCIENCE*

PHYSICS

Majorana fermions go for a ride

Evidence for propagating Majorana quasiparticles is found in a topological superconductor

By **Sumanta Tewari**[1] *and* **Tudor D. Stanescu**[2]

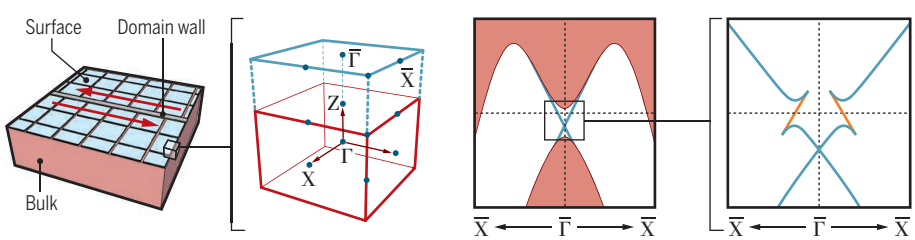
Enrico Fermi described Ettore Majorana as having the mind of a genius. The Majorana fermion, born as a testimony to the truthfulness of mathematical aesthetics, has recently returned to the center stage of modern physics. These are particles that are also their own antiparticles. For decades, Majorana's theory was considered a mathematical curiosity that has little to do with reality. However, they have now become the key concept associated with certain types of quasiparticles in condensed-matter systems (*1*). In the condensed-matter context, they are not fundamental particles like electrons or neutrinos but emerging excitations that we term quasiparticles. On page 104 of this issue, Wang *et al.* (*2*) provide strong evidence for the observation of Majorana quasiparticles in an iron-based superconductor, $FeSe_xTe_{1-x}$.

The concept of Majorana fermions has its roots in a celebrated equation discovered in the late 1920s by physicist Paul Dirac. The Dirac equation seamlessly brings together quantum mechanics and the special theory of relativity and provides the quantum mechanical description of spin-half fermions, such as electrons, protons, and neutrons. The equation has the property that if a solution exists with an energy $+E$, $-E$ is also a solution. The positive-energy solutions describe the "regular" particles (for example, electrons and neutrons), whereas the negative-energy solutions—initially regarded by even Dirac as unphysical—are now known to describe their so-called "antiparticles" (for example, positrons for electrons and antineutrons for neutrons). For every quantum mechanical particle that obeys the Dirac equation, there are subatomic antiparticles with physical properties that are identical in some respects to those of the corresponding particle (for example, the same mass) and exactly opposite in some others (for example, baryon numbers and opposite electric charges). Thus, the negatively charged electron has an antielectron called a positron that is the positively charged counterpart with equal mass. The neutron and the antineutron have the same mass but opposite baryon numbers.

In the late 1930s, Ettore Majorana showed that the Dirac equation accepts a class of solutions that describes particles that are identical in every respect to their antiparticles. The original candidates for Majorana fermions—neutrons and neutrinos—appeared to have distinct antiparticles. For neutrinos, the jury is still out, but the concept of Majorana fermions has become central to developments in supersymmetry, dark matter, and, most recently, certain types of emerging quasiparticles in condensed-matter systems. In condensed-matter systems, they can emerge in a special class of supercon-

Inferring Majorana modes

Indirect observations support the existence of propagating Majorana quasiparticles (QPs) in the presence of surface topological superconducting states (blue) emerging from the bulk states (red).



Setting the stage
The iron-based topological superconductor has a domain wall (black line) separating two regions shifted by one-half of a unit cell.

Defining the directions
The reciprocal lattice points of the bulk (X, Γ, Z) and surface states ($\bar{X}$, $\bar{\Gamma}$) are shown.

Topological surface states
In the bulk, valence and conducting bands (red) do not touch. Surface valence and conduction bands (blue) touch at one point.

Majorana QPs
Along the domain wall, a linear set of allowed low-energy states (orange) would connect the gapped bands of the topological superconductor.

Published by AAAS

ductors called topological superconductors as either quasiparticles localized at certain types of defects, such as vortices, or as delocalized quasiparticles propagating along the boundaries or the walls between different domains of superconductivity (see the figure). In condensed-matter physics, the Majorana quasiparticles are predicted to obey fundamentally new particle statistics that generalize from the Fermi-Dirac statistics governing identical particles with half-integer spin. Experimental observations of sharp tunneling conductance peaks that are consistent with the presence of Majorana quasiparticles in semiconductor-superconductor heterostructures (*3*–*5*) and ferromagnetic atomic chains deposited on superconductors (*6*) have been reported in the past few years, although definitive evidence is still lacking (*7*). These observations have only been on static quasiparticles, otherwise called localized or bound states. By contrast, evidence for Majorana quasiparticles that propagate along the boundaries or walls between different domains in superconductors has not been observed.

"Experimental observations... are consistent with the presence of Majorana quasiparticles..."

The challenge associated with the experimental demonstration of Majorana quasiparticles stems, essentially, from the complexity of the proposed host systems. These systems are multicomponent heterostructures (*3*–*6*) predicted to behave as topological superconductors under controlled external conditions. By contrast, Wang *et al.* exploit a key advantage of the iron-based superconductor $FeSe_xTe_{1-x}$ in that it consists of a single material that supports all the key ingredients necessary for Majorana physics. This requires a nontrivial electronic band structure, superconductivity, and special types of domain walls that can support propagating Majorana quasiparticles. The surface of this system is a topological superconductor with a special band structure, based on an argument put forward by Fu and Kane (*8*). A domain wall separating regions of the crystalline lattice shifted by half a unit-cell should support a pair of counterpropagating Majorana quasiparticles identifiable with a linear energy-momentum dispersion relation (*8*).

Wang *et al.* used a combination of previous studies and experimental and theoretical evidence to show that this scenario exists for $FeSe_xTe_{1-x}$. Previous studies identified topological superconducting surface states with high-resolution angle-resolved photoemission spectroscopy (*9*) and observed sharp zero-bias peaks inside vortex cores (*10*). The authors' scanning tunneling microscopy studies show a flat (bias-independent) differential conductance along the domain wall. This signature is the hallmark of the linearly dispersing Majorana quasiparticles propagating along a one-dimensional defect. The spatial distribution of domain wall states as a function of energy is also consistent with the evolution expected from the Majorana quasiparticle being localized along the domain wall at zero energy to become delocalized as its energy approaches the gap edge. The authors also show a robust zero bias anomaly observed in vortex cores below the superconducting transition temperature T_c, which is a signature of companion Majorana quasiparticles localized at vortex cores consistent with previous studies (*10*).

The experiments from Wang *et al.* provide a compelling case for topological superconductivity and propagating Majorana quasiparticles in a class of iron-based superconductors. This system is attractive because the observations and theory suggest Majorana behavior occur in a single material, without a heterostructure. The findings open a new chapter in the field of iron-based superconductors and represent a large step in the quest for Majorana fermions in condensed-matter systems. For practical applications, Majorana systems in condensed-matter physics are attractive for quantum computing because their fundamentally new particle statistics may be useful for the development of fault-tolerant topological quantum computation (*11*, *12*). This makes these iron-based systems important both for fundamental science and quantum technology. ■

REFERENCES AND NOTES

1. N. Read, D. Green, *Phys. Rev. B* **61**, 10267 (2000).
2. Z. Wang *et al.*, *Science* **367**, 104 (2020).
3. J. D. Sau, R. M. Lutchyn, S. Tewari, S. Das Sarma, *Phys. Rev. Lett.* **104**, 040502 (2010).
4. V. Mourik *et al.*, *Science* **336**, 1003 (2012).
5. H. Zhang *et al.*, *Nature* **556**, 74 (2018).
6. S. Nadj-Perge *et al.*, *Science* **346**, 602 (2014).
7. C. Moore, C. Zeng, T. D. Stanescu, S. Tewari, *Phys. Rev. B* **98**, 155314 (2018).
8. L. Fu, C. L. Kane, *Phys. Rev. Lett.* **100**, 096407 (2008).
9. P. Zhang *et al.*, *Science* **360**, 182 (2018).
10. D. Wang *et al.*, *Science* **362**, 333 (2018).
11. A. Kitaev, *Phys. Uspekhi* **44**, 131 (2001).
12. C. Nayak, S. H. Simon, A. Stern, M. Freedman, S. Das Sarma, *Rev. Mod. Phys.* **80**, 1083 (2008).

[1]Department of Physics and Astronomy, Clemson University, Clemson, SC 29634, USA. [2]Department of Physics and Astronomy, West Virginia University, Morgantown, WV 26506, USA. Email: stewari@clemson.edu

10.1126/science.aaz6961

DEVELOPMENTAL BIOLOGY

Building a carnivorous trap

Experiments and computations reveal developmental origins of cup-shaped leaves

By **Derek E. Moulton** *and* **Alain Goriely**

Variation, according to evolutionary biologist Stephen Jay Gould, is "nature's only irreducible essence" (*1*). The variation and diversity of shapes in nature is a central focus of both evolutionary and developmental biologists. Unified under the unlikely roof of "evolutionary developmental biology," the ultimate goal of these scientists is to understand how variation arises both through natural selection (on geological time scales) and during development (on embryological time scales). On page 91 of this issue, Whitewoods *et al.* (*2*) present a fascinating example of evolutionary developmental biology in a carnivorous plant.

The herbarium of our early school years taught us that leaves come in many different sizes, shapes, and textures that have evolved by subtle gene rearrangements to solve various packing and arrangement problems (*3*). Whereas many leaves are nearly flat, so as to present their best face to the Sun, others have been sculpted by evolution to function as mechanical devices; ropes, springs, spikes, claws, spears, hooks, catapults, and traps are the medieval weapons that plants use in their daily struggle with the environment. Perhaps most intriguing are carnivorous traps, the revenge of the plant kingdom and one of the lesser-known interests of Darwin (*4*).

The humped bladderwort (*Utricularia gibba*), is an inconspicuous, easy-to-grow aquatic plant found on all inhabited continents. Yet, it has developed a nearly spherical cage and a sophisticated release mechanism that can swallow an unsuspecting crustacean in a few milliseconds (*5*). How can a leaf develop into such an elegant and complicated structure?

With its small genome, the bladderwort turns out to be an excellent model system. In these plants, the same branch supports both needle-like leaves and bladder-shaped traps

Oxford Centre for Industrial and Applied Mathematics (OCIAM), Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK. Email: goriely@maths.ox.ac.uk

Published by AAAS

(see the figure). Both structures develop from the same initial dome-shaped organ primordia. The selection of one shape or the other occurs through the physical translation of a delicate variation in differential gene expression and morphogen production, which can be probed experimentally and theoretically through computational modeling.

During its development, the leaf must solve a problem in geometry. There is a fundamental difference between a flat sheet and a sphere, a fact that can be appreciated by trying to flatten an orange peel. The two states are geometrically incompatible, so transforming one into the other involves either stretching or cutting. Because the integrity of the leaf is preserved through development, the incompatible morphological change from dome to needle or trap can occur only through differential growth that can be either anisotropic (varied rates of growth in different directions) or heterogeneous (varied rates of growth in different parts of the organ). For example, a small sphere of material may be deformed into an ellipsoid or with a bulge on one side (see the figure).

The challenge, then, is to decipher the genetic underpinnings of the necessary growth differentials. Much is known about the effects of particular genes on the shape of leaves (*6*). In developing leaves, key genes are expressed differentially in zones on the adaxial (upper) versus abaxial (lower) surfaces. Whitewoods *et al.* revealed that these same genes are expressed differently in leaflets that form traps versus ones that form needle-like leaves. This crucial observation was confirmed by showing that trap development can be inhibited by inducing expression of one of the genes in an abnormal position on a leaflet.

Having established relevant gene expression profiles, Whitewoods *et al.* turned to a computational model for leaf morphogenesis. Most computational work in plant biology tends to model morphology by tracking cell growth and division (*7*). One of the singular features of the authors' research was their study of growth deformations at the tissue level. In their model, each point in the budding organ—treated as a three-dimensional continuum—was given different growth rates in each of three selected directions. These directions are linked to a polarity field obtained by the diffusion of a morphogen. Their model is an adaptation of the theory of morphoelasticity (*8*), which allows for continuous changes resulting from both mechanics and growth. This theory has been used successfully in animal morphogenesis to describe the formation of a wide range of structures, from folds in the brain to seashell architecture (*9*, *10*).

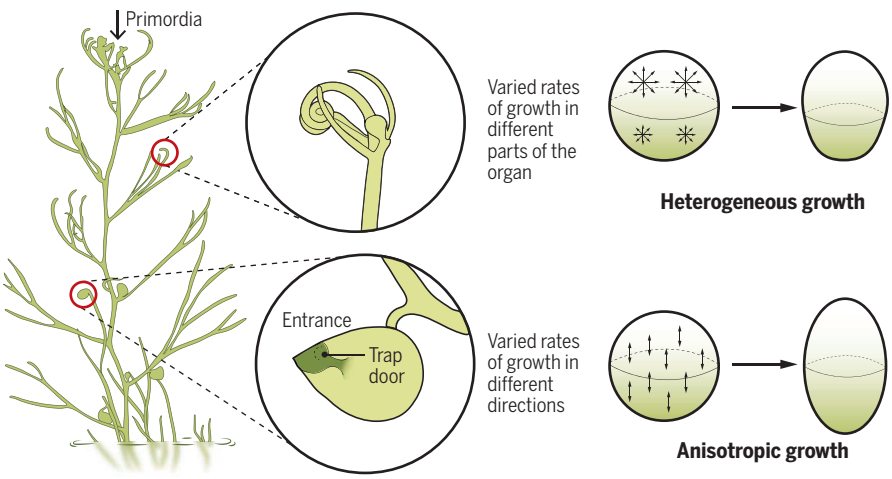
Notably, by simply varying the growth rates, the computational model showed how the same dome can develop into either a needle-like cone or a planar leaf shape. By making these growth rates nonuniform in space, Whitewoods *et al.* further demonstrated a simple mechanism for generating cup-shaped traps and other features, such as the ridges found in a related cousin, *Sarracenia purpurea* (*11*). The model also provided some insight into a chicken-and-egg problem in development: Does the orientation of cell division generate growth anisotropy, or does growth anisotropy generate the orientation of cell division? The new model demonstrates a logical sequence of differential gene expression preceding growth polarity, which precedes cell-division orientation.

By showing that spherical traps, conical needles, and flat leaves all can be generated from the same initial tissue shape through small shifts in gene expression and growth differentials inspired by morphogen distribution, the new study opens exciting lines of research. For example, can these polarity fields be explicitly identified by measurements of gene expression at the cellular level? How are the polarity fields influenced by chemical and mechanical stimuli (*12*)? And how exactly does the developmental process form a functioning trap? The trap mechanism involves the slow build-up and rapid release of mechanical energy (*13*), which is intimately linked to morphological changes during development (*14*), but the connection has not yet been explored.

More broadly, Whitewoods *et al.* offer key insights into the competing pressures that ultimately shape every living thing. Development is inherently a physical process and is thus the end result of physical forces subtly manipulated by genetic clues. To understand such a process requires analysis across multiple scales as well as the integrated tools of mechanics, mathematics, and biology (*15*). However, this multidisciplinary approach tells only half the story of evolutionary developmental biology. On the species scale, evolution is driven by forces that enable one organism to successfully reproduce while another dies out. To properly connect evolutionary and embryonic forces across vastly different scales is to understand the very nature of variation: "the hard reality" (*1*). ■

Bladderwort beginnings

During development of the humped bladderwort, the same initial primordia (arrow) on a single branch can be transformed into either needle-like leaves (left, top circle) or carnivorous traps (left, bottom circle). The key to shaping an organ is differential gene expression, which creates differential growth that is either heterogeneous or anisotropic.



GRAPHIC: N. DESAI/*SCIENCE*

REFERENCES AND NOTES

1. S. J. Gould, *Discover* **6**, 40 (1985).
2. C. D. Whitewoods *et al.*, **367**, 91 (2020).
3. D. Kierzkowski *et al.*, *Cell* **177**, 1405 (2019).
4. C. Darwin, *Q. J. Microsc. Sci.* **1874**, 185 (1874).
5. O. Berg, M. D. Brown, M. J. Schwaner, M. R. Hall, U. K. Müller, *J. Exp. Zool. A Ecol. Integr. Physiol.* **333**, 29 (2020).
6. E. E. Kuchen *et al.*, *Annu. Rev. Plant Biol.* **62**, 365 (2012).
7. P. Prusinkiewicz, A. Runions, *New Phytol.* **193**, 549 (2012).
8. A. Goriely, "*The Mathematics and Mechanics of Biological Growth*" (Springer, 2017).
9. D. Ambrosi *et al.*, *J. R. Soc. Interface* **16**, 20190233 (2019).
10. D. E. Moulton, A. Goriely, R. Chirat, *Proc. Natl. Acad. Sci. U.S.A.* 10.1073/pnas.1916520116 (2019).
11. K. Fukushima *et al.*, *Nat. Commun.* **6**, 6450 (2015).
12. O. Hamant *et al.*, *Science* **322**, 1650 (2008).
13. O. Berg, K. Singh, M. R. Hall, M. J. Schwaner, U. K. Müller, *Integr. Comp. Biol.* **2019**, icz144 (2019).
14. H. Hofhuis *et al.*, *Cell* **166**, 222 (2016).
15. V. Mirabet, P. Das, A. Boudaoud, O. Hamant, *Annu. Rev. Plant Biol.* **62**, 365 (2011).

ACKNOWLEDGMENTS

A.G. is supported by the Engineering and Physical Sciences Research Council, grant EP/R020205/1.

10.1126/science.aba3797

Published by AAAS

FLUORESCENT PROTEINS

Electrostatics affect the glow

Chromophore twisting is probed with unnatural amino acids

By Cheng Hu[1], Xiaohong Liu[2], Jiangyun Wang[2]

The chromophores of fluorescent proteins (FPs) form through self-catalyzed posttranslational modifications (*1*). In the original green FP (GFP) isolated from the jellyfish *Aequorea victoria*, Ser[65], Tyr[66], and Gly[67] residues form the 4-(*p*-hydroxybenzylidene)-5-imidazolinone (HBI) chromophore that contains a phenolate ring (P-ring), an imidazoline ring (I-ring), and a monomethine bridge (*1*). The protein cage excludes water that can quench fluorescence, but also enhances the fluorescence quantum yield (FQY) by restricting bond-twisting photoisomerization of the HBI chromophore. However, the protein could also improve FQY through electrostatic effects. As reported on page 76 of this issue, Romei *et al.* (*2*) studied the effect of introducing groups that donate or withdraw chromophore electrons on the FQY of the photoswitchable FP Dronpa2 (*3*).

Mutations of residues near the chromophore can fine-tune the bond-rotation energy barrier to create photoswitchable variants. In Dronpa2, green light produces the protonated fluorescent cis form, and blue light produces the nonprotonated nonfluorescent trans form. This switching enables applications in super-resolution imaging (*4*) and optogenetics (*5*). In this protein, the chromophore is formed from Cys[62], Tyr[63], and Gly[64]. Investigation of the contribution of the electrostatics to FQY requires fine-tuning electrostatic parameters in the complex protein environment. Genetic code expansion (*6*) allows unnatural amino acids (UAAs) to be introduced into FPs. Yu *et al.* (*7*) and others established methods for the genetic incorporation of large numbers of Tyr analogs to modify the chromophore.

Upon photon absorption, the chromophore enters the S_1 state, in which negative charge flows from the P-ring to the I-ring (see the figure). Romei *et al.* found that replacing Tyr[63] with analogs bearing an electron-withdrawing group required more energy to transfer the electron, which blue-shifted the absorption maximum, and substitution with an electron-donating group red-shifted the absorption maximum. Both substitutions decreased FQY and likely lowered the energy barrier for bond rotation.

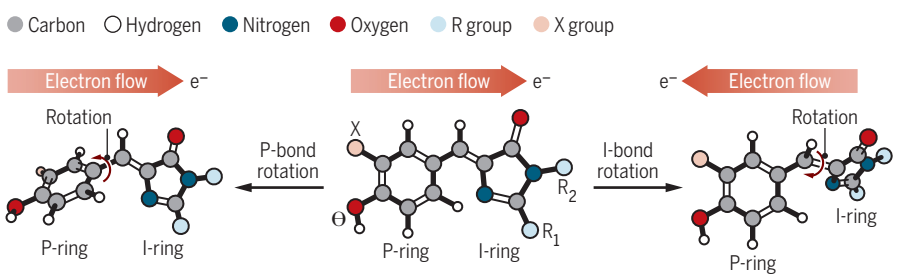
The next challenge was to decipher the mechanism of how bond twisting is modulated by electrostatics (*8*). In the transformation from the S_1 state to the fluorescence off-state 1, a cis-trans isomerization of the double bond connecting the P-ring and I-ring is required. Previous studies of the photoswitchable FP rsEGFP2 showed that in the transition state, the P-ring and I-ring are nearly perpendicular (*9*). This geometry reduces the amount of π-π orbital overlap between the P-ring and I-ring, which results in negative charge flowing back from the I-ring to the P-ring. Alternatively, in the S_1 state, the P-ring can rotate to reach another fluorescence off-state 2, and this process causes more negative charge flows from the P-ring to the I-ring. Electron-donating group substitution on Tyr[63] can lower the excited-state barrier for P-bond rotation and decreases FQY.

Romei *et al.* show that the electric field exerted by the chromophore's environment can either promote or hinder charge transfer, and thereby could effectively control the choice of bond rotation and isomerization pathway after photoexcitation. For optogenetic applications, photoinduced rotation around a specific bond of the FP chromophore or a retinal chromophore triggers a distinct conformational change of the protein, which results in specific kinase activation (*5*) or ion conductance (*10*). Engineering the electrostatic and steric environment of the chromophore by introducing charged, hydrogen-bonding, or unnatural amino acids in the protein scaffold could lead to more precise control of chromophore twisting and downstream signaling pathways, or create photoswitchable FPs with higher total photon number and faster on-off state switching for super-resolution imaging (*4*).

Recent advances in time-resolved serial femtosecond crystallography with x-ray free electron lasers (XFELs) have visualized bond-rotation events in proteins (*9*). Transient absorption spectroscopy could probe how genetically encoded UAAs influence the I-bond rotation and P-bond rotation pathways. These insights could in turn inform new mutagenesis efforts to develop improved FPs. ■

Excited-state outcomes

Romei *et al.* used genetic code expansion of a Tyr residue to change the H atom (denoted X) of a protein chromophore to groups that withdraw electrons (such as Cl) or donate electrons (such as OCH_3) to explore the effects of electrostatics on fluorescence (R_1 and R_2 are contacts to the protein).



To off-state 2
Rotation around the P-ring leads to an off-state and moves more charge to the I-ring.

On state (cis)
Photoexcitation creates the cis fluorescent S_1 state. Charge flows from the P-ring to the I-ring.

To off-state 1 (trans)
Crossing the bond-rotation barrier creates the trans off-state. Charge flows back to the P-ring.

[1]Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. [2]Institute of Biophysics, Chinese Academy of Sciences, Chaoyang District, Beijing, China. Email: jwang@ibp.ac.cn

REFERENCES AND NOTES

1. A. Acharya *et al.*, *Chem. Rev.* **117**, 758 (2017).
2. M. G. Romei, C.-Y. Lin, I. I. Mathews, S. G. Boxer, *Science* **367**, 76 (2020).
3. A. C. Stiel *et al.*, *Biochem. J.* **402**, 35 (2007).
4. Z. Fu *et al.*, *Nat. Methods* 10.1038/s41592-019-0613-6 (2019).
5. X. X. Zhou, L. Z. Fan, P. Li, K. Shen, M. Z. Lin, *Science* **355**, 836 (2017).
6. J. W. Chin, *Nature* **550**, 53 (2017).
7. Y. Yu, X. Liu, J. Wang, *Acc. Chem. Res.* **52**, 557 (2019).
8. C. Y. Lin, J. Both, K. Do, S. G. Boxer, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E2146 (2017).
9. N. Coquelle *et al.*, *Nat. Chem.* **10**, 31 (2018).
10. K. K. Yang, Z. Wu, F. H. Arnold, *Nat. Methods* **16**, 687 (2019).

ACKNOWLEDGMENTS

Supported by the National Science Foundation of China (21750003, 21837005, 21890743, and U1632133) and the Chinese Academy of Sciences (QYZDB-SSW-SMC032).

10.1126/science.aba0571

GRAPHIC: A. KITTERMAN/*SCIENCE*

Published by AAAS

PHOTO: STRINGER/REUTERS/NEWSCOM

Research is needed to understand the projected increase in cancer incidence and mortality in sub-Saharan Africa.

GLOBAL HEALTH

Cancer in sub-Saharan Africa

Knowledge of cancer in Africa brings needed diversity to improve health worldwide

By **Timothy R. Rebbeck**

Cancer is an increasing global public health burden. This is especially the case in sub-Saharan Africa (SSA); high rates of cancer—particularly of the prostate, breast, and cervix—characterize cancer in most countries in SSA. The number of these cancers in SSA is predicted to more than double in the next 20 years (*1*). Both the explanations for these increasing rates and the solutions to address this cancer epidemic require SSA-specific data and approaches. The histopathologic and demographic features of these tumors differ from those in high-income countries (HICs). Basic knowledge of the epidemiology, clinical features, and molecular characteristics of cancers in SSA is needed to build prevention and treatment tools that will address the future cancer burden. The distinct distribution and determinants of cancer in SSA provide an opportunity to generate knowledge about cancer risk factors, genomics, and opportunities for prevention and treatment globally, not only in Africa.

The most frequent cancers in African countries include prostate, lung, liver, leukemia, non-Hodgkin's lymphoma, and Kaposi's sarcoma in men and breast and cervical cancer in women (see the figure). Distinct risk factors in SSA contribute to the understanding of cancer etiology in ways that may not be as easily studied in HICs. For example, the observation of high rates of Burkitt's lymphoma in SSA launched the research that identified Epstein-Barr virus as the causal agent of Burkitt's lymphoma (*2*), thus providing some of the earliest knowledge about the infectious and molecular etiology of cancer. Similarly, the high prevalence of HIV infection and the attendant large number of HIV-associated cancers in SSA have enabled research that has led to understanding of the causes and treatment of cancers exacerbated by HIV infection, including cervical cancer, Kaposi's sarcoma, and non-Hodgkin's lymphoma (*3*).

Patients with cancer in SSA are often diagnosed when their disease is in advanced stages. This is in part a consequence of inadequate resources for cancer prevention and early detection. Delayed diagnosis coupled with inadequate treatment options is a major reason for the continent's cancer mortality rates, which are 1.5- to 4-fold higher than in HICs for leading cancers (*1*). Studies that focus on features of late-stage and aggressive disease will be required to better understand how to manage such disease states. This research is required to develop and implement early detection and treatment modalities that can be implemented in low-resource settings such as SSA, where availability of laboratory or imaging technologies is limited.

Although molecular and other biological data that address cancer etiology and progression in SSA are limited, emerging evidence suggests that distinct tumor histopathology, tumor subtypes, and molecular signatures exist in SSA. For example, Nigerian breast cancer cases were defined by increased mutational signature associated with deficiency of the homologous recombination DNA repair pathway, pervasive mutations in the tumor suppressor gene *TP53*, mutations in GATA binding protein 3 (*GATA3*), and greater genomic mutational burden (indicating aggressive biology), compared with breast tumors from African Americans or Caucasians (*4*). This suggests different tumor etiologies by race or geography, perhaps reflecting particular environmental exposures to carcinogens, and may provide knowledge about the spectrum of breast cancer molecular phenotypes that may be composed of distinct molecular subtypes and represent different frequencies of known subtypes in non-Africans. Knowledge of these differences can lead to optimized monitoring and treatment across all populations.

Early age at diagnosis is a hallmark of cancer in SSA, which has been proposed to reflect a higher rate of hereditary cancer in SSA. This is supported by the observation that although 29% of SSA women and 33% of Caucasian women are between ages 25 and 49, 58% of SSA women are diagnosed with breast cancer before they are 50 years old, compared with 21% of Caucasian women (*5*). Breast tumors in SSA Black cases are twice as likely as SSA Caucasian cases to be the triple-negative subtype, meaning that the estrogen, progesterone, and HER2 receptors are not expressed by the breast cancer cells, making them highly aggressive and difficult to treat (*6*). These molecular features are hallmarks of hereditary cancers; genetic testing for pathogenic sequence variants in the tumor suppressor genes breast cancer 1 (*BRCA1*) and *BRCA2* in a series of Nigerian breast cancer cases suggested that the rate of cancers in women with *BRCA1* or *BRCA2* germline pathogenic sequence variants is higher than in Caucasian populations (*7*). Although large population studies of hereditary cancer in SSA do not yet exist, *BRCA1* and *BRCA2* mutations of SSA origin are also found in African Americans, and the type of mutations in *BRCA1* and *BRCA2* differs between SSA and non-SSA populations (*8*).

Generation of knowledge about cancer in SSA will lead to both improved cancer prevention and treatment. Protocols and networks are being formed to study cancer in SSA patients and translate this knowledge to health care in SSA (*9*). These protocols may begin with those used in HICs but must be optimized to low-resource settings, which have limited access to equipment, trained personnel, and therapies. Although the availability of basic research infrastructure is more limited in SSA than in HICs, there are numerous basic science and translational research institutes in SSA, including those that can address

Dana-Farber Cancer Institute and Harvard T. H. Chan School of Public Health, Boston, MA, USA.
Email: timothy_rebbeck@dfci.harvard.edu

Published by AAAS

cancer genomics and molecular biology. Pan-African initiatives, such as the Human Heredity and Health in Africa (H3Africa) initiative, have developed common protocols, data systems, and collaborative networks that are developing the infrastructure and resources required to address cancer needs in SSA. Africa-based organizations that foster cancer research, translation, infrastructure, and training also exist, including the African Academy of Sciences and the African Organization for Research and Training in Cancer. The World Economic Forum has initiated a "Leapfrogging with Precision Medicine" initiative that will advance the use of genetics and genomics in cancer prevention and treatment. These and many other institutions and activities have the potential to develop the knowledge and sustainable resources needed to address the cancer burden in SSA.

In addition to benefiting the treatment of cancer patients in SSA, the knowledge gained from research on the continent will be informative for cancer globally. Evidence suggests that diversity in study populations will improve the ability to generate and generalize knowledge that can be applied to cancer and other diseases. For example, rare mutations judged to be pathogenic on the basis of Caucasian mutation frequency data were later found to occur commonly in African American populations (*10*). Knowledge of mutation frequency subsequently led to the opposite conclusion that these variants were nonpathogenic. This misclassification of pathogenicity led to molecular misdiagnoses that could have been avoided had diverse populations been included in the original study cohort. Similarly, it has been observed that use of artificial intelligence and other "big data" approaches in diagnostic and therapeutic applications will be suboptimal if these approaches are trained and validated in nondiverse populations.

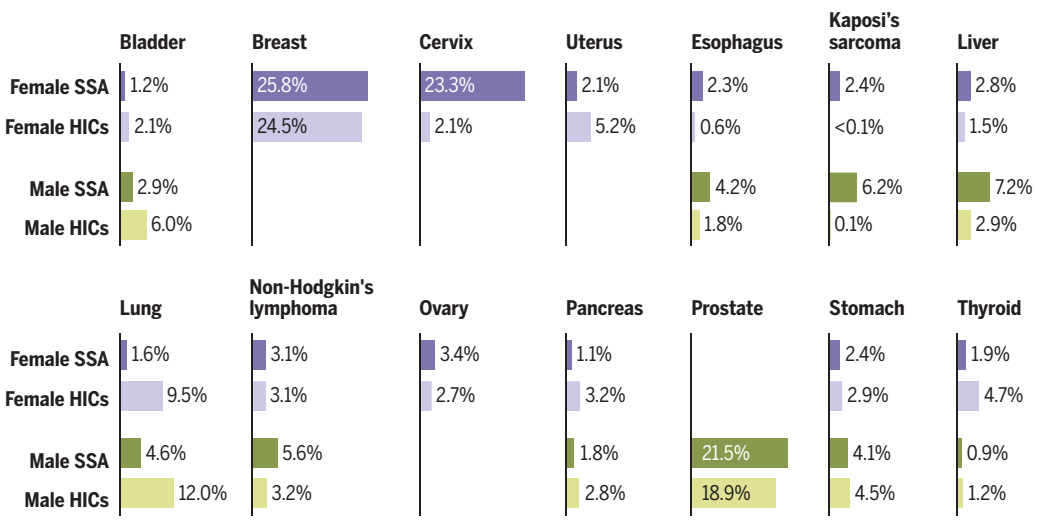
A limited scope of population data may perpetuate or exacerbate existing biases in the information needed to generate effective clinical and public health interventions (*11*). Thus, it is not only the underrepresented population that will benefit from increased research representation but all populations. There is growing evidence that genetic risk prediction models developed in Caucasian populations may not be appropriate in African-descent populations. These results suggest that ancestry-specific models may be required to optimally predict cancer risk. Such models may use genetic variants defined from the discovery of relevant race-specific disease-associated variants and/or use ancestry-specific genomic markers to define ancestry rather than self-identified race or ethnicity.

Because of the underlying genetic and genomic relationships between Africans and members of the African diaspora (primarily in North America and Europe), knowledge gained from research in SSA can be used to address health disparities that are prevalent in members of the African diaspora. West African genomic ancestry (the ancestral origin of most African Americans) has been reported to confer the highest genomic risk for prostate cancer of any population worldwide (*12*). Selection against a prostate cancer susceptibility locus on chromosome 2q37 has led to lower frequencies of this risk allele in Caucasians and higher frequencies in Africans and African Americans (*12*). KhoeSan (an indigenous southern African people) ethnicity has also been associated with prostate cancer risk (*13*). These data from African populations may help elucidate the causes of cancer disparities, particularly in African Americans.

Diagnostic, monitoring, and treatment technologies developed out of necessity in low-resource settings provide an opportunity for cost-efficient and accessible technologies that can be implemented in HICs. In particular, development of point-of-care technologies that provide rapid and accurate cancer diagnosis or treatment in SSA could be applied in community settings to address disparities in access to cancer services in HICs. When technologies, including data-derived models of risk or prognosis, are not developed so that they can be applied in diverse settings and populations, health disparities can be created or increased (*14*). Development of technologies that can be used in SSA could provide knowledge about optimal implementation of these technologies worldwide. Advances in science and technology and the concomitant increase in capacity for health improvements will increase diversity as a means to remove bias from cancer genomics studies and improve the quality of cancer prevention and treatment globally. ■

Cancer in Africa compared with high-income countries

The occurrence of the common types of cancer in sub-Saharan Africa (SSA) differs from that in high-income countries (HICs). The newly diagnosed cancer cases in 2018, as a percentage of the total population, are shown for SSA versus HICs. It is estimated that incidence of the most common cancers will approximately double between 2018 and 2040 in SSA. Data were obtained from (*15*).



GRAPHIC: X. LIU/*SCIENCE*

REFERENCES AND NOTES

1. D. Parkin *et al.*, Eds., *Cancer in Sub-Saharan Africa* (IARC Scientific Publication no. 167, IARC, 2018).
2. D. Burkitt, *Br. J. Surg.* **46**, 218 (1958).
3. L. Chinula *et al.*, *Curr. Opin. HIV AIDS* **12**, 89 (2017).
4. J. J. Pitt *et al.*, *Nat. Commun.* **9**, 4181 (2018).
5. Population Division, United Nations Department of Economic and Social Affairs, *2019 Revision of World Population Prospects*; https://population.un.org/wpp.
6. V. A. McCormack *et al.*, *Breast Cancer Res.* **15**, R84 (2013).
7. Y. Zheng *et al.*, *J. Clin. Oncol.* **36**, 2820 (2018).
8. T. R. Rebbeck *et al.*, *Hum. Mutat.* **39**, 593 (2018).
9. C. Andrews *et al.*, *J. Glob. Oncol.* **4**, 1 (2018).
10. A. K. Manrai *et al.*, *N. Engl. J. Med.* **375**, 655 (2016).
11. G. Kuhnen, A. Rebhan (2019); www.advisory.com/daily-briefing/2019/02/04/ai.
12. J. Lachance *et al.*, *Cancer Res.* **78**, 2432 (2018).
13. D. C. Petersen *et al.*, *BMC Med. Genomics* **12**, 82 (2019).
14. D. Weiss *et al.*, *PLOS ONE* **13**, e0195447 (2018).
15. The Global Cancer Observatory; https://gco.iarc.fr.

ACKNOWLEDGMENTS

The author is supported by National Institutes of Health grant U01-CA184734.

10.1126/science.aay4743

Published by AAAS

CELL BIOLOGY

Gene expression regulated by RNA stability

The factor responsible for autoregulation of tubulin RNA stability is identified

By **Ofer Shoshani** *and* **Don W. Cleveland**

One of the first discoveries of gene expression mediated by controlling messenger RNA (mRNA) stability is autoregulation of tubulin synthesis. In this regulatory process, the concentration of tubulin subunits modulates the stability of the mRNAs from which they are translated (*1*, *2*). In the 1980s it was found that only translated tubulin mRNAs are autoregulated (*3*) and that translation had to continue through at least 41 amino acids (*4*). This is enough for the nascent tubulin polypeptide to emerge from the ribosome (*5*). Later work established that when the tubulin subunit pool is high, the first four amino acids (Met-Arg-Glu-Ile, MREI) emerging during nascent tubulin translation serve as a regulatory tag. Recognition of this tag promoted the degradation of the translating tubulin mRNA (*4*, *6*–*8*). More than 30 years later, on page 100 of this issue, Lin *et al.* (*9*) identify tetratricopeptide repeat protein 5 (TTC5) as the regulator that binds to nascent tubulin polypeptides.

With the exception of a high-resolution confirmation that an elevated pool of tubulin subunits selectively represses tubulin synthesis (*8*), no progress toward understanding the autoregulation of tubulin expression was made since 1988. The most attractive model for how the pool size of tubulin subunits could trigger rapid mRNA degradation to suppress new tubulin synthesis was that it was the tubulin α/β dimer, the unit that assembles into microtubules, that cotranslationally bound to the MREI tetrapeptide. Lin *et al.* disprove this model. They use mass spectrometry and in vitro translation of an mRNA encoding the first 94 amino acids of β-tubulin to identify TTC5 as the protein that recognizes the nascent β-tubulin MREI tetrapeptide in complex with the large ribosomal subunit.

α- and β-tubulin form a heterodimer that serves as the building block for microtubule polymers, the tracks along which cargoes are moved by dynein and kinesin family motors. During cell duplication, microtubule-directed trafficking is essential for delivery to each daughter cell of a complete set of chromosomes. By inactivating both maternal and paternal copies of the gene encoding TTC5, Lin *et al.* demonstrate that tubulin autoregulation is essential for maintaining faithful chromosome segregation, with a modest increase in chromosome segregation errors in the absence of TTC5. Errors in chromosome inheritance are key drivers of tumorigenesis, so maintenance of the genome is important (*10*). However, Lin *et al.* have determined that cells with inactivated tubulin autoregulation are viable and can continue to survive and duplicate. This is unexpected because disruption of autoregulation would be predicted to yield runaway tubulin synthesis, which in turn should have severely disrupted microtubule assembly dynamics and function. Perhaps additional factors are involved, the activities of which might compensate, to some extent, for the lack of TTC5 activity.

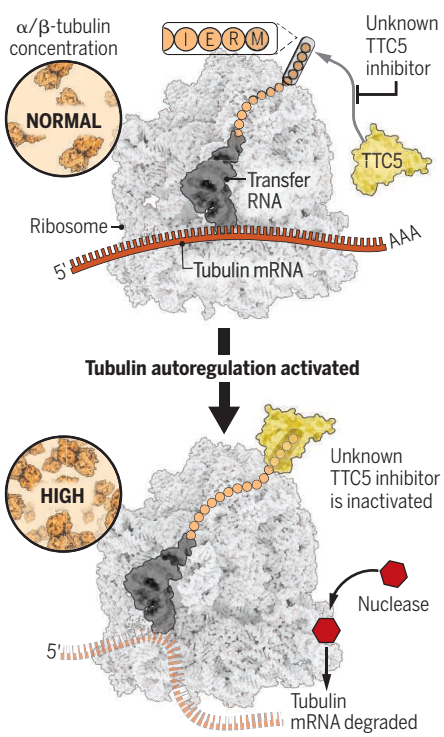
The new work also casts doubt on the notion that the tubulin concentration directly participates in the ribosomal complex with TTC5 and the MREI peptide. Instead, the findings of Lin *et al.* support a counterintuitive model in which cells ordinarily contain a cytosolic factor that prevents TTC5 binding to MREI-ribosome complexes and that this inhibitor is inactivated when the tubulin dimer concentration increases (see the figure).

The study of Lin *et al.* is a major step in deciphering a regulatory pathway for controlling expression of an important cellular product (tubulin) through cotranslationally mediated mRNA instability. It should be noted, however, that important steps in the autoregulatory pathway remain unidentified, including (i) the factor that inhibits TTC5 binding activity when tubulin levels are normal, (ii) the newly proposed autoregulatory signal generated by an increased pool of tubulin subunits, (iii) the nuclease(s) that mediate tubulin mRNA degradation, and (iv) the activation of those nucleases once TTC5 recognizes the nascent tubulin peptide.

This may only be the tip of the iceberg for cotranslational control of gene expression. Work in yeast suggests that mRNA decay through cotranslational regulation is widespread (*11*) and that it involves a 5′-to-3′ RNA exoribonuclease 1 (xrn1) (*12*). The TTC5 structure bound to a ribosome, as beautifully determined by Lin *et al.*, provides clues that might allow the identification of additional regulators (perhaps containing the same tetratricopeptide repeats found in TTC5) that can bind the ribosome exit tunnel and simultaneously recognize targets encoded by different mRNAs. ■

Autoregulation of tubulin RNA

Tubulin messenger RNA (mRNA) instability is mediated by cotranslational binding of tetratricopeptide repeat protein 5 (TTC5) to the tubulin amino-terminal tetrapeptide MREI and activation of one or more ribosome-associated nucleases.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

Ludwig Institute for Cancer Research and Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA 92093, USA. Email: dcleveland@ucsd.edu

REFERENCES AND NOTES

1. T. J. Yen, D. A. Gay, J. S. Pachter, D. W. Cleveland, *Mol. Cell. Biol.* **8**, 1224 (1988).
2. D. W. Cleveland, M. A. Lopata, P. Sherline, M. W. Kirschner, *Cell* **25**, 537 (1981).
3. D. W. Cleveland, M. F. Pittenger, J. R. Feramisco, *Nature* **305**, 738 (1983).
4. D. A. Gay, T. J. Yen, J. T. Lau, D. W. Cleveland, *Cell* **50**, 671 (1987).
5. J. S. Pachter, T. J. Yen, D. W. Cleveland, *Cell* **51**, 283 (1987).
6. T. J. Yen, P. S. Machlin, D. W. Cleveland, *Nature* **334**, 580 (1988).
7. A. Yonath, K. R. Leonard, H. G. Wittmann, *Science* **236**, 813 (1987).
8. I. Gasic, S. A. Boswell, T. J. Mitchison, *PLOS Biol.* **17**, e3000225 (2019).
9. Z. Lin *et al.*, *Science* **367**, 100 (2020).
10. U. Ben-David, A. Amon, *Nat. Rev. Genet.* **21**, 44 (2020).
11. V. Pelechano, W. Wei, L. M. Steinmetz, *Cell* **161**, 1400 (2015).
12. W. Hu, T. J. Sweet, S. Chamnongpol, K. E. Baker, J. Coller, *Nature* **461**, 225 (2009).

10.1126/science.aba0713

Published by AAAS

POLICY FORUM

ENERGY

Sustainable minerals and metals for a low-carbon future

Policy coordination is needed for global supply chains

By **Benjamin K. Sovacool[1], Saleem H. Ali[2,3,4], Morgan Bazilian[5], Ben Radley[6], Benoit Nemery[7], Julia Okatz[8], Dustin Mulvaney[9]**

Climate change mitigation will create new natural resource and supply chain opportunities and dilemmas, because substantial amounts of raw materials will be required to build new low-carbon energy devices and infrastructure (*1*). However, despite attempts at improved governance and better corporate management, procurement of many mineral and metal resources occurs in areas generally acknowledged for mismanagement, remains environmentally capricious, and, in some cases, is a source of conflict at the sites of resource extraction (*2*). These extractive and smelting industries have thus left a legacy in many parts of the world of environmental degradation, adverse impacts to public health, marginalized communities and workers, and biodiversity damage. We identify key sustainability challenges with practices used in industries that will supply the metals and minerals—including cobalt, copper, lithium, cadmium, and rare earth elements (REEs)—needed for technologies such as solar photovoltaics, batteries, electric vehicle (EV) motors, wind turbines, fuel cells, and nuclear reactors. We then propose four holistic recommendations to make mining and metal processing more sustainable and just and to make the mining and extractive industries more efficient and resilient.

Between 2015 and 2050, the global EV stock needs to jump from 1.2 million light-duty passenger cars to 965 million passenger cars, battery storage capacity needs to climb from 0.5 gigawatt-hour (GWh) to 12,380 GWh, and the amount of installed solar photovoltaic capacity must rise from 223 GW to more than 7100 GW (*3*). The materials and metals demanded by a low-carbon economy will be immense (*4*). One recent assessment concluded that expected demand for 14 metals—such as copper, cobalt, nickel, and lithium—central to the manufacturing of renewable energy, EV, fuel cell, and storage technologies will grow substantially in the next few decades (*5*). Another study projected increases in demand for materials between 2015 and 2060 of 87,000% for EV batteries, 1000% for wind power, and 3000% for solar cells and photovoltaics (*6*). Although they are only projections and subject to uncertainty, the World Bank put it concisely that "the clean energy transition will be significantly mineral intensive" (*7*) (see the figure).

Many of the minerals and metals needed for low-carbon technologies are considered "critical raw materials" or "technologically critical elements," terms meant to capture the fact that they are not only of strategic or economic importance but also at higher risk of supply shortage or price volatility (*8*). But their mining can produce grave social risks. A majority of the world's cobalt, used in the most common battery chemistries for EVs and stationary electricity storage, is mined in the Democratic Republic of Congo (DRC) (see the map), a country struggling to recover from years of armed conflict. There, women and sometimes children often work in or around mines for less pay or status than their male and adult counterparts, without basic safety equipment (see the photo). Owing to a lack of preventative strategies and measures such as drilling with water and proper exhaust ventilation, many cobalt miners have extremely high levels of toxic metals in their body and are at risk of developing respiratory illness, heart disease, or cancer.

In addition, mining frequently results in severe environmental impacts and community dislocation. Moreover, metal production itself is energy intensive and difficult to decarbonize. Mining for copper, needed for electric wires and circuits and thin-film solar cells, and mining for lithium, used in batteries, has been criticized in Chile for depleting local groundwater resources across the Atacama Desert, destroying fragile ecosystems, and converting meadows and lagoons into salt flats. The extraction, crushing, refining, and processing of cadmium, a by-product of zinc mining, into compounds for rechargeable nickel cadmium batteries and thin-film photovoltaic modules that use cadmium telluride (CdTe) or cadmium sulfide semiconductors can pose risks such as groundwater or food contamination or worker exposure to hazardous chemicals, especially in the supply chains where elemental cadmium exposures are greatest. REEs, such as neodymium and the less common dysprosium, are needed for magnets in electric generators in wind turbines and motors in EVs, control rods for nuclear reactors, and the fluid catalysts for shale gas fracking. But REE extraction in China has resulted in chemical pollution from ammonium sulfate and ammonium chloride and tailings pollution that now threaten rural groundwater aquifers as well as rivers and streams. Several metals for green technologies are found as "companions" to other ores with differential value and unsustainable supply chains (*9*).

POLICY RECOMMENDATIONS

With these sobering social and environmental aspects of current mineral extraction in mind, we suggest four policy recommendations.

Diversify mining enterprises for local ownership and livelihood dividends

Although large-scale mining is often economically efficient, it has limited employment potential, only set to worsen with the recent arrival of fully automated mines. Mining can concentrate occupational hazards as well as environmental risk, as demonstrated most severely by tailings pond disasters and mining wastewater contamination. Even where there is relative political stability and stricter regulatory regimes in place, there can still be serious environmental failures, as exemplified by the recent global rise in dam failures at settling ponds for mine tailings. The level of distrust of extractive industries has even led to countrywide moratoria on all new mining projects, such as in El Salvador and the Philippines.

Traditional labor-intensive mechanisms of mining that are possible to undertake with less mechanization and without major capital investments are called artisanal and small-scale mining (ASM). Although ASM is not immune from poor governance or environmental harm, it provides livelihood potential for at least 40 million people worldwide, with an additional three to five times more people indirectly supported by the sector (*10*). It is also usually more strongly embedded in local and national economies than foreign-owned, large-scale mining, with a greater level of

[1]University of Sussex, Brighton, UK. [2]University of Delaware, Newark, DE, USA. [3]University of Queensland, Brisbane, Queensland, Australia. [4]United Nations International Resource Panel, United Nations Environment Programme, Nairobi, Kenya. [5]Colorado School of Mines, Golden, CO, USA. [6]London School of Economics, London, UK. [7]Katholieke Universiteit Leuven, Leuven, Belgium. [8]SYSTEMIQ Ltd., London, UK. [9]San José State University, San José, CA, USA. Email: b.sovacool@sussex.ac.uk

Published by AAAS

value retained and distributed within the country. Diversifying mineral supply chains to allow for greater coexistence of small- and large-scale operations is needed. Yet, efforts to incorporate artisanal miners into the formal economy have often resulted in a scarcity of permits awarded, exorbitant costs for miners to legalize their operations, and extremely lengthy and bureaucratic processes for registration.

Development donors need to focus on bottom-up formalization efforts rather than merely facilitating government efforts to better regulate the sector for increased tax revenues. There needs to be a focus on policies that recognize its livelihood potential in areas of extreme poverty. Moreover, formalization of the sector should focus on creating stronger, more accountable arrangements to drive greater value of resource revenues down the supply chain to ASM miners to ensure better environmental and safety mechanisms and expand their access to markets. The recent decision of the London Metals Exchange to have a policy of "nondiscrimination" toward ASM is a positive sign in this regard. Certain industry actors have demonstrated a commitment to, and the benefits of, this type of approach, such as Fairphone's sourcing of the mineral columbite-tantalite (coltan) used in mobile phones. At the level of government policy, ASM has demonstrated its ability to increase productivity and mechanize production, even in hostile regulatory and governance environments. More space for and support to ASM to pursue this trajectory would enhance its capacity to meet the increased demand for minerals required in the move toward a low-carbon future. One place to begin is with the redistribution of dormant mining concessions previously granted to (but unused by) mining companies so that local ASM operators can legally work in these locations, as has been taking place recently in Tanzania.

A creuseur, or digger, descends into a Congolese copper and cobalt mine in Kawama. Wages are low, and working conditions are dangerous, often with no safety equipment or structural support for the tunnels.

PHOTO: MICHAEL ROBINSON CHAVEZ/*THE WASHINGTON POST* VIA GETTY IMAGES

Acknowledge the limits of traceability

A great deal of attention has focused on fostering transparency and accountability of mineral mining by means of voluntary traceability or even "ethical minerals" schemes. International groups, including Amnesty International, the United Nations, and the Organisation for Economic Co-operation and Development, have all called on mining companies to ensure that supply chains are not sourced from mines that involve illegal labor and/or child labor. In concert, Eurasian Resources Group (ERG) launched their Clean Cobalt Framework in 2018, First Cobalt has their Responsible Cobalt Initiative, RCS Global has its Better Cobalt program, Amnesty International is working on an Ethical Battery framework, and the World Economic Forum launched a Global Battery Alliance committed to "responsible sourcing" of raw materials for batteries.

Traceability schemes, however, may be impossible to fully enforce in practice and could, in the extreme, merely become an exercise in public relations rather than improved governance and outcomes for miners. In the eastern DRC, for example, cassiterite, the mineral that tin is extracted from, is exported through a traceability system yet can nonetheless have contributed to conflict financing or labor and human rights abuses while simultaneously introducing heavy financial costs onto local workers for the right to participate in the system (*11*). Nonetheless, traceability is not without promise, and examples from Blockchain technology show how the use of artificial intelligence algorithms for data processing has the potential for greater assurance but ultimately relies on the accuracy of data being fed into the supply chain.

Transparency of supply chains is a means to an end and will only be effective if consumers or regulators start to differentiate between products being provided. There are effective lessons on traceability and transparency arising from the Kimberley Process for conflict diamonds; the Extractive Industries Transparency Initiative for oil, gas, and mineral resources; and the Fairmined Standard for gold that could be applied to the mineral supply chains needed for decarbonization. Paramount among these is an acknowledgment that traceability schemes offer a largely technical solution to profoundly political problems and that these political issues cannot be circumvented or ignored if meaningful solutions for workers are to be found. Traceability schemes ultimately will have value if the market and consumers trust their authenticity and there are few potential opportunities for leakage in the system.

Explore new resource streams

Although primary emphasis must be placed on resource efficiency (higher output or usage of product per unit of resource input) and recycling, there will likely be a need for primary resource extraction as well owing to clean-energy infrastructure demand. New resource streams—including metal availability in seawater (desalination) and groundwater (geothermal brines), material substitution or material intensity reductions, and materials recovery and recycling—also hold promise for diversifying supply chains, as long as they maintain environmental sustainability and protect worker safety.

Although mining in terrestrial areas is likely to continue to meet the demands of low-carbon technologies in the nearer term, we need to carefully consider mineral sources beneath the oceans in the longer term. The

Published by AAAS

International Seabed Authority, set up under the United Nations (UN) Convention on the Law of the Sea, is in the process of issuing regulations related to oceanic mineral extraction. This process is a rare opportunity to be proactive in setting forth science-based environmental safeguards for mineral extraction. For metals such as cobalt and nickel, ocean minerals hold important prospects on the continental shelf within states' exclusive economic zones as well as the outer continental shelf regions. Within international waters, metallic nodules found in the vast Clarion-Clipperton Zone of the Pacific as well as in cobalt and tellurium crusts, which are found in seamounts worldwide, provide some of the richest deposits of metals for green technologies. Difficult extraction and declining reserves of some terrestrial minerals, as well as social resistance against terrestrial mining, may lead to oceanic mineral reserves becoming more plausible sources. Minerals near hydrothermal vents are in more pristine and distinctive ecosystems and should likely remain off-limits for mineral extraction for the foreseeable future.

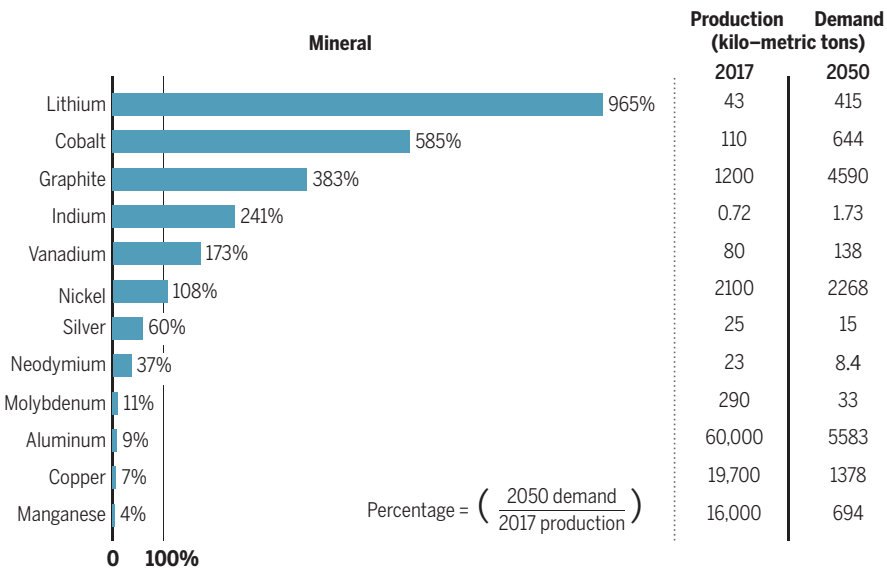
Technological substitution can play an important role as well. Copper offers an illustrative example. Higher copper prices in recent years have incentivized replacement in new applications in the automotive industry, such as wire harnesses and replacing copper with aluminum winding in motors. However, substitution to other primary metals or even synthetics could merely shift resource demand to another material that may be more abundant initially but can become more challenging to procure over time. Moreover, substitution may be limited to particular innovations or niches. Alternatives to lithium-ion batteries, such as sodium-ion batteries, are becoming more practical and feasible. But finding substitutes for metals like platinum group metals in key technologies such as fuel cells has become increasingly difficult, and reserves are dwindling.

Recycling and better resource efficiency can play a part at extending and enhancing the lifetimes of products and also stretching out mineral reserves. Closed-loop supply chains based on circular economy ideas in addition to advancements in metallurgy, reverse logistics, waste separation, materials science, waste processing, and advanced recycling can all enhance the longevity and continual reuse of minerals and metals. Researchers at the U.S. National Renewable Energy Laboratory estimate that 65% of the U.S. domestic cobalt demand in 2040 could be supplied by end-of-life lithium-ion batteries, provided a robust take-back and recycling infrastructure is in place.

Extended producer responsibility (EPR) is a framework that stipulates that producers are responsible for the entire lifespan of a product, including at the end of its usefulness. EPR would, in particular, shift responsibility for collecting the valuable resource streams and materials inside used electronics from users or waste managers to the companies that produce the devices. EPR holds producers responsible for their products at the end of their useful life and encourages durability, extended product lifetimes, and designs that are easy to reuse, repair, or recover materials from. A successful EPR program known as PV Cycle has been in place in Europe for photovoltaics for about a decade and has helped drive a new market in used photovoltaics that has seen 30,000 metric tons of material recycled. To date, EPR has mainly shaped collection, recycling, and waste management to ensure safe and responsible disposal of specific classes of products like e-waste, paint, and pharmaceuticals, but, in concept, it is also meant to help drive more sustainable design as well as options for reuse and repair. There is evidence of EPR's influence on green design in the global solar industry. For example, thin-film manufacturer First Solar screens new materials to ensure that they will not negatively influence their recycling process, through which they currently recover 90% of their CdTe semiconductor material and 90% of their glass. To more easily recycle the plastics and copper from photovoltaics, some manufacturers are seeking out halogen-free components.

Space mining, although potentially useful for developing lunar and planetary bases farther into the future, has less potential for meeting the demand for minerals for immediate decarbonization on Earth. A possible exception to this may be platinum group metals from asteroids, but here, too, the time frame and quantity of production would preclude its use in meeting immediate technology needs for climate mitigation.

Growth in mineral needs for low-carbon energy technology



All production and demand data reflect annual values. 2017 data reflect annual production for all uses. 2050 data reflect estimated demand for only low-carbon energy technology uses. Data from (*7*).

Incorporate minerals into climate and energy planning

Given the centrality of minerals and metals to the future diffusion of low-carbon technologies, materials security should be actively incorporated into formal climate planning. This could be connected to ongoing planning as part of the nationally determined contributions (NDCs) under the Paris Accord, the European Commission's National Energy and Climate Plans (NECPs), or even energy policy-making at the national scale. Climate planners could begin by mapping out their NDC contributions alongside a list of "critical minerals" for energy security (see supplementary materials).

Although care must be taken to ensure that the NDC process does not become too broad or research intensive, we believe the NDCs are the most tangible international policy consensus mechanism on this matter. The NDCs can incorporate some of the mineral sourcing challenges through efforts at resource efficiency. The Group of Seven (G7) has taken on this linkage, and policies to motivate resource efficiency can be a means of keeping track of material

GRAPHIC: X. LU/*SCIENCE*

Published by AAAS

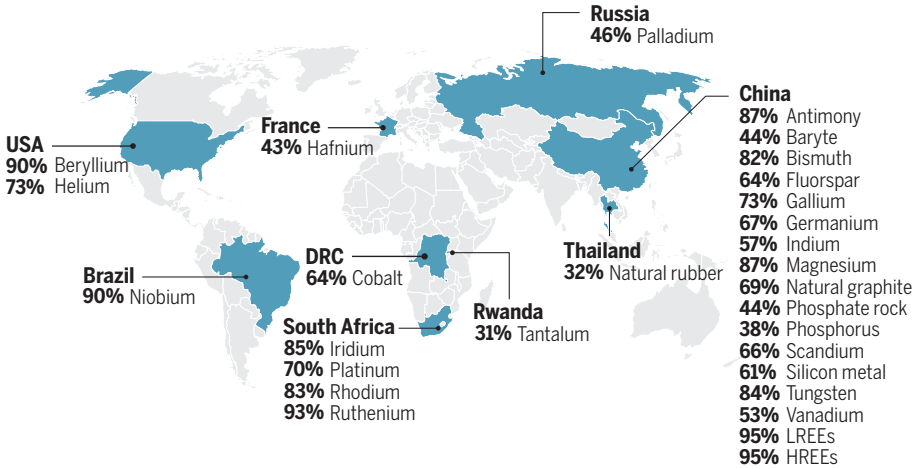
and mineral supply chains. For example, a materials assessment for particular infrastructure options for climate change mitigation or adaptation could be included in cost-benefit analyses. Recent work has suggested that the social acceptability of tying resource-efficient products to climate change mitigation efforts is strong (*12*).

Having each country create a list of critical minerals within its NDC process and show possible trade-offs and shortfalls could lead to several benefits. More efforts on national critical material analysis could result in improved mapping of mineral supply chains, for which there is already a notable gap across many developing countries and regions. The analytical efforts would enhance our understanding of supply constraints and demand patterns, which in turn could lead to a better understanding of future prices and drivers, especially those beyond the control of governments and policy as agents of change. The process of mapping mineral demands for NDCs, NECPs, and national energy policies could lead to new linkages and networks and a raising of awareness, connecting the traditional minerals and metals community to other research and social communities, especially in climate policy and energy studies. In this way, climate mitigation could be twinned with minerals security and industrial strategy as a way to meet broad sets of goals (environmental, political, and economic) in one stroke.

AN ETHICAL CONUNDRUM

Mineral and metal supplies are geologically determined, yet socially mediated. Even if supplies are enhanced through coproducts of other industries, new resource streams, and considerable expansion of recycling and increased recovery rates, there are likely to be bottlenecks across metal supply chains (*13*). This is exacerbated by poorly functioning markets, as least for the minor metals. Hence, trade policy will need to become more deftly aligned with mineral supply in ways which are both economically and ecologically more efficient. Furthermore, more robust reporting and emissions data will be required across the supply chain. For example, although the U.S. government strategy for mineral supply security released in June 2019 highlights the importance of trade with allies and partners, it does not consider where it is most ecologically efficient to source minerals. Pursuing decarbonization simultaneously with principles of a circular economy, coupled with increased market transparency mechanisms and full lifecycle reporting, could yield important social and environmental benefits.

Countries accounting for the largest share of critical raw materials



DRC, Democratic Republic of Congo; LREEs, light rare earth elements; HREEs, heavy rare earth elements. Figure modified from European Commission, "Third list of critical raw materials for the EU of 2017" (European Commission, 2017); https://ec.europa.eu/growth/sectors/raw-materials/specific-interest/critical_en.

Consideration should also be given to where mining is most likely to have a positive development footprint while also having more manageable environmental impacts (*14*). Utilizing tools such as the Responsible Mining Index and platforms such as the Responsible Minerals Initiative or the Intergovernmental Forum on Mining, Minerals, Metals and Sustainable Development may be a way forward. Although there may be treaty fatigue among policy-makers, an intertreaty protocol on mineral supply chains to ensure that the goals of existing treaties are met could enhance effective governance. Conversations in this vein should be attempted among the parties to the UN Framework Convention on Climate Change, through the UN Environment Assembly, as well as more focused mechanisms such as the U.S. government's recently launched Energy Resource Governance Initiative, the World Bank's Climate-Smart Mining Facility, or the European Institute of Innovation and Technology for Raw Materials.

Having just marked the 150th anniversary of the formulation of the periodic table, it is high time we realize that the elements, and the minerals in which they are embedded, are essential to our attainment of low-carbon goals. There is an ethical conundrum to addressing climate change only by aggravating other social and ecological problems related to unsustainable mineral and metal supply chains. But done sustainably, an impending mining boom could help lift communities out of poverty, accelerate technical innovation for decarbonization, and further the realization of energy and climate targets. Which direction it takes will depend considerably on how metal and mineral supply chains are governed over the next few critical years. ■

REFERENCES AND NOTES

1. Intergovernmental Panel on Climate Change (IPCC), "Global warming of 1.5°C," V. Masson-Delmotte *et al.*, Eds. (IPCC Special Report, IPCC, 2018); www.ipcc.ch/sr15.
2. D. M. Franks *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 7576 (2014).
3. International Renewable Energy Agency (IRENA), "Global energy transformation: A roadmap to 2050" (IRENA, Abu Dhabi, 2018).
4. International Resource Panel, "Global resources outlook 2019: Natural resources for the future we want" (United Nations Environment Programme, Nairobi, 2019); www.resourcepanel.org/reports/global-resources-outlook.
5. E. Dominish, N. Florin, S. Teske, "Responsible minerals sourcing for renewable energy" (Institute for Sustainable Futures, University of Technology, Sydney, 2019).
6. A. Månberger, B. Stenqvist, *Energy Policy* **119**, 226 (2018).
7. World Bank, "Climate-smart mining: Minerals for climate action" (World Bank, 2018); www.worldbank.org/en/topic/extractiveindustries/brief/climate-smart-mining-minerals-for-climate-action.
8. European Commission, "Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee, and the Committee of the Regions on the 2017 list of critical raw materials for the EU" (COM/2017/490, European Commission, Brussels, 2017).
9. N. T. Nassar, T. E. Graedel, E. M. Harper, *Sci. Adv.* **1**, e1400180 (2015).
10. World Bank, "2019 state of the artisanal and small-scale mining sector" (World Bank, Washington, DC, 2019).
11. B. Radley, C. Vogel, *Extr. Ind. Soc.* **2**, 406 (2015).
12. C. E. Cherry, K. Scott, J. Barrett, N. F. Pidgeon, *Nat. Clim. Chang.* **8**, 1007 (2018).
13. S. H. Ali *et al.*, *Nature* **543**, 367 (2017).
14. L. J. Sonter, S. H. Ali, J. E. M. Watson, *Proc. R. Soc. London Ser. B* **285**, 20181926 (2018).

ACKNOWLEDGMENTS

This project received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 730403 "Innovation pathways, strategies and policies for the low-carbon transition in Europe (INNOPATHS)." The content of this deliverable does not reflect the official opinion of the European Union. Responsibility for the information and views expressed herein lies entirely with the author(s).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6473/30/suppl/DC1

10.1126/science.aaz6003

GRAPHIC: ADAPTED FROM EUROPEAN UNION/CC-BY BY X. LU/*SCIENCE*

Published by AAAS

NEXTGEN VOICES

Making science accessible

To mark the 30th anniversary of the Americans with Disabilities Act, we asked young scientists this question: **What one thing would you change about the training or careers in your field to improve accessibility for people with visible and/or invisible disabilities?** A selection of their responses is below. Follow NextGen Voices on Twitter with hashtag #NextGenSci. Read previous NextGen Voices survey results at https://science.sciencemag.org/collection/nextgen-voices.

Provide logistical support

Because I have dyslexia, I use read-out-loud software, but it is not well suited to journal articles with citations. Listening to citations makes keeping track of the paper difficult and can add hours of reading time. Some expensive software allows you to select what to read, but it can take days to mark the citations and is impractical when reading many journal articles. A tool that allows the listener to skip the citations would help.

Kristen Tuosto
Center for the Advanced Study of Human Paleobiology, The George Washington University, Washington, DC 20052, USA.
Email: ktuosto@gwu.edu

Movies have subtitles for a reason. People might be hearing impaired or easily distracted, have trouble deciphering accents, or understand written information better. Speech-to-text technology, which enables instructors to instantly provide written versions of their lecture materials, should be used in real time during classroom discussions. Alternatively, written summaries prepared beforehand should accompany visual and oral presentations. This would allow learners to read the material as explained by the professor, rather than a textbook, multiple times for stronger comprehension.

Juliet Tegan Johnston
Department of Civil, Environmental, and Geo-Engineering, University of Minnesota, Minneapolis, MN 55455, USA. Twitter: @queermsfrizzle

Wet labs are designed for people with no physical challenges. Many of us who could contribute to science have difficulties not with the techniques or knowledge but rather with the need to spend hours sitting on a lab stool. Adjustable benches and equipment would improve accessibility and the wet lab experience for all scientists.

Carol Connolly
Dublin 18, Ireland.
Email: Hateyes@yahoo.com

Many fields of medicine require physicians to work on call for more than 24 hours. Although this serves as important experience, students with disabilities may find it prohibitively challenging to work so many consecutive hours. A reduction in required on-call hours would allow students with and without disabilities to thrive in their training. Students pursuing specialties with extended-call commitments could elect to gain this exposure.

Cody Lo
Faculty of Medicine, University of British Columbia, Vancouver, BC V6T 1Z2, Canada. Twitter: @cody_lo

Despite being the most popular event at scientific conferences, poster sessions are probably the least accessible. As someone who struggles with low vision, I find poster sessions frustrating. Most posters have unreadable fonts, unclear images, and poor color contrast. Conference organizers should encourage poster presenters to make a 2-minute video summary of their work. The video and the presenter's contact information could then be accessed by scanning a QR code. These QR codes would help build camaraderie among researchers; unlike business cards, conference goers would have details of the presenters together with their presentations.

Edmond Sanganyado
Provincial Key Laboratory of Marine Biotechnology, Shantou University, Shantou, Guangdong, 515063, China. Twitter: @ESanganyado

In physical anthropology, the majority of assessments of ancestry, sex, and age of a skeleton are based on morphological features. However, only a limited number of models are provided to demonstrate the differences. The inclusion of 3D models with differing variations of features would be helpful to every student, not just the ones who are differently abled. I believe it would bridge the divide between recently graduated students and experienced practitioners, allowing for greater mentorship of other technical aspects of the work.

Kristy A. Winter
College of Humanities, Arts, and Social Sciences, Flinders University, Bedford Park, SA 5042, Australia. Twitter: @Winterhiccup

Having a disability can be very expensive. In addition to costs such as out-of-pocket fees for therapy, people with mobility impairments may pay more for rent, as accessible apartments tend to be more expensive, or more for food because of specific dietary needs. These higher costs contribute to a higher risk of poverty for people with disabilities. Therefore, increasing pay for Ph.D. students would make it more likely that students with a disability—as well as students from

ILLUSTRATION: ROBERT NEUBECKER

Published by AAAS

disadvantaged backgrounds—could pursue a scientific career.

Tanja Roembke
Institute of Psychology, RWTH Aachen University, Aachen, NRW 52062, Germany.
Email: tanja.roembke@psych.rwth-aachen.de

Much of the scientific work in the field of chemistry depends on colors. For those who are color blind, the use of techniques such as acid-base titrations or spectroscopy presents a challenge. Technology could help. If an app could analyze a smartphone camera photo and correct different types of color blindness by converting the colors into decipherable patterns, it could be used not only in the lab but also to convert figures in textbooks and papers into more accessible images.

Wagner Eduardo Richter
Department of Chemical Engineering, Technological Federal University–Paraná, Paraná, Brazil. Email: richter@utfpr.edu.br

Hearing impairment affects a sizeable minority of world inhabitants. To better welcome all scientists, we must incorporate sign-language translation and closed-captioning services in scientific meetings and conferences. Computer technologies are advancing, and machine translation services are now available at a fraction of their former cost. This simple modification could have a lasting impact on the progression of science through inclusion.

Kyle J. Isaacson
Department of Biomedical Engineering, University of Utah, Salt Lake City, UT 84112, USA.
Twitter: @kjisaacson

Foster inclusive community

When onboarding a new member, the research group leader should affirm the entire group's commitment to accommodating anyone with a disability. Affirming this commitment to every new member ensures people with invisible disabilities are not unintentionally ignored, empowers people with a disability to disclose information and request accommodations, and sets clear expectations for those without disabilities to make accommodation requests a priority. Having this conversation during onboarding also helps form a community that is supportive of people with disabilities and in turn benefits from their unique perspectives.

Michael Raitor
Department of Mechanical Engineering, Stanford University, Stanford, CA 94305, USA.
Email: mraitor@stanford.edu

As a researcher with cerebral palsy, I know first-hand the fear that comes with applying for a job when you have a disability. Will department leaders be willing to make accommodations, and will they worry that hiring me will affect productivity? Academia needs to avoid direct comparison between researchers with and without disabilities. The academic paths of researchers with disabilities may differ, for example, because of prolonged stays in hospitals or numerous surgeries and rehabilitations. However, these "hardships" make us better researchers: They push us to develop excellent organizational skills, creativity, stress resilience, emotional intelligence, and grit. Students in STEM fields need role models in the form of lecturers or professors with disabilities. They need to see that it can be done and that universities and departments are inclusive in reality and not only on paper.

Aleksandra Kosanic
Ecology Team, Department of Biology, University of Konstanz, D 78457 Konstanz, Germany. Twitter: @SashaKosanic

In my field of astronaut training and operations, it is clear that humans sometimes need equipment to thrive in their environment. Just as an astronaut needs specialized technology to survive in space and a diver needs scuba gear to survive in water, people with disabilities benefit from access to the right tools. Technology, whether a brain-controlled exoskeleton, artificial limbs, or a computer-generated voice, allows us to overcome environmental challenges and transform into superhumans. If more people understood this concept, more individuals with disabilities would be involved in scientific fields.

Loredana Bessone
European Space Agency, Cologne, NRW 51147, Germany. Twitter: @esa_caves

In graduate school, mental health is often put on the back burner. Encouraging more discussions about and providing more institutional support for mental health would foster a more welcoming and accepting environment. It would also allow all graduate students the opportunity to develop a better work-life balance and become more productive scientists.

Ashley B. Heim
School of Biological Sciences, University of Northern Colorado, Greeley, CO 80639, USA.
Email: ashley.heim@unco.edu

The shame and stigma of disabilities caused by depression and post-traumatic stress disorder prevent the people who have them from asking for help. Breaking this barrier would be easier if colleagues and especially principal investigators and group leaders were trained to identify the signs and reach out to ensure their support. Building the basic bridge of trust and safety is the most important first step in accessing people with hidden disabilities. Supervisors should be trained to create and communicate a safe, kind environment and to access further resources for help.

Pragya Srivastava
Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104, USA.
Twitter: @Prags0211

Facilitating accessible learning spaces for students with invisible disabilities poses unique challenges, since neither student nor instructor is initially aware of the other's condition. Moreover, some invisible diagnoses, such as autism spectrum disorder (ASD), carry a degree of social stigma. If they feel comfortable doing so, I believe it would be valuable for course instructors and graduate mentors affected by ASD to acknowledge this fact to their students. This would make it easier for students to open up about their own invisible challenges and reduce both perceived and actual stigmatization.

P. William Hughes
Department of Ecology, Environment, and Plant Sciences, Stockholm University, Stockholm, 10691, Sweden. Twitter: @pwilliamhughes

We need to change the narrative to value the experience of overcoming adversity rather than viewing those who have overcome challenges as inferior or damaged. I'd love to see a seminar series where established principal investigators could openly discuss the challenges they faced and how they overcame those challenges. This would increase representation and a sense of belonging among people with invisible disabilities who might not otherwise know who else is out there. It would also reduce stereotyping from able scientists.

Caitlin M. Aamodt
Neuroscience Interdepartmental PhD Program, University of California, Los Angeles, Los Angeles, CA 90025-4816, USA. Twitter: @CaitlinAamodt

10.1126/science.aba6129

ILLUSTRATION: ROBERT NEUBECKER

Published by AAAS

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

Light micrograph showing cortical neurons with branching dendrites

NEUROSCIENCE

Human dendrites are special

A special developmental program in the human brain drives the disproportionate thickening of cortical layer 2/3. This suggests that the expansion of layer 2/3, along with its numerous neurons and their large dendrites, may contribute to what makes us human. Gidon *et al.* thus investigated the dendritic physiology of layer 2/3 pyramidal neurons in slices taken from surgically resected brain tissue in epilepsy patients. Dual somatodendritic recordings revealed previously unknown classes of action potentials in the dendrites of these neurons, which make their activity far more complex than has been previously thought. These action potentials allow single neurons to solve two long-standing computational problems in neuroscience that were considered to require multilayer neural networks. —PRS

Science, this issue p. 83

MULTIPLEX GENOMICS

Single-cell chemical transcriptomics

Single-cell transcriptomic technologies have emerged as powerful tools to explore cellular heterogeneity at the resolution of individual cells. Srivatsan *et al*. now add another layer of information and complexity by combining single-cell transcriptomics with oligo hashing and small molecule screening in a method called sci-Plex. Because screening many chemical compounds requires the ability to profile many cells, and because screens perturb cells in many different ways, the authors demonstrate the effects of 188 compounds in three cancer lines. The sci-Plex method can capture gene expression profiles from thousands of experimental conditions in a single experiment. —LMZ

Science, this issue p. 45

NANOPARTICLES

Finding the fivefold path

Crystals formed in nature or synthesized in a lab can have domains with different orientations that are called twins. Having five of these domains is relatively common, but how these fivefold twins form is still a mystery. Song *et al*. combined in situ transmission electron microscope observations with simulations to find two different pathways for forming fivefold twins in gold, platinum, and palladium nanoparticles. This combination provides a key insight into the twin-forming process, which might be exploited for developing nanoparticles suitable for a variety of applications. —BG

Science, this issue p. 40

APPLIED PHYSICS

It's knot what you know

Why is it that some knots seem to hold tight while others readily slip apart? Patil *et al*. develop a theoretical analysis of the stability of knots and find links between topological parameters (twist charge, crossing numbers, handedness) and mechanical stability. The theory is confirmed using simulations and experiments on color-changing fibers that optically show localized stress differences in different parts of the knot as the two strands are pulled apart. The authors show why some common knots slip easily and untie, whereas others hold tight. —MSL

Science, this issue p. 71

OPTICS

Miniaturizing particle accelerators

Particle accelerators are usually associated with large national facilities. Because photons are able to impart momentum to electrons, there are also efforts to develop laser-based particle accelerators. Sapra *et al*. developed an integrated particle accelerator using photonic inverse design methods to optimize the interaction between

PHOTO: CNRI/SCIENCE SOURCE

Published by AAAS

the light and the electrons. They show that an additional kick of around 0.9 kilo–electron volts (keV) can be given to a bunch of 80-keV electrons along just 30 micrometers of a specially designed channel. Such miniaturized dielectric laser accelerators could open up particle physics to a number of scientific disciplines. —ISO

Science, this issue p. 79

ARCHAEOLOGY

Middle Stone Age cooking

Early evidence of cooked starchy plant food is sparse, yet the consumption of starchy roots is likely to have been a key innovation in the human diet. Wadley *et al*. report the identification of whole, charred rhizomes of plants of the genus *Hypoxis* from Border Cave, South Africa, dated up to 170,000 years ago. These archaeobotanical remains represent the earliest direct evidence for the cooking of underground storage organs. The edible *Hypoxis* rhizomes appear to have been cooked and consumed in the cave by the Middle Stone Age humans at the site. *Hypoxis* has a wide geographical distribution, suggesting that the rhizomes could have been a ready and reliable carbohydrate source for *Homo sapiens* in Africa, perhaps facilitating the mobility of human populations. —AMS

Science, this issue p. 87

EVOLUTIONARY BIOLOGY

Generating a new transcriptional network

Organismal novelties result from changes in transcriptional circuits. But what comes first, changes in regulatory protein or changes in cis-regulatory sequences? Britton *et al*. examine the Matα2 protein in a Saccharomycotina clade of fungi. They show that a newly evolved transcription circuit involving repression of the a-specific genes by the ancient homeodomain protein Matα2 occurred in two stages separated by millions of years. In the first stage, Matα2 acquired several coding changes followed by changes in cis-regulatory sequences. This clade-specific requirement explains how the coding changes of Matα2 were in place long before the new a-specific gene repression circuit arose. —BAP

Science, this issue p. 96

PARASITIC INFECTIONS

Parasite perturbation of immunity

Helminths—a broad grouping of parasitic worms—infect billions of people and are known to modulate host immune responses to promote their own survival. De Ruiter *et al*. used mass cytometry to determine which immune cells are affected by helminth infection. They analyzed samples from rural Indonesians before and after deworming treatment. These were compared to samples from Europeans and urban Indonesians, neither of which had been exposed to helminths. Helminths expanded specific type 2 and regulatory immune cells, and many immune alterations reverted upon deworming. These details on host–pathogen interaction could inform future targeted therapies. —LP

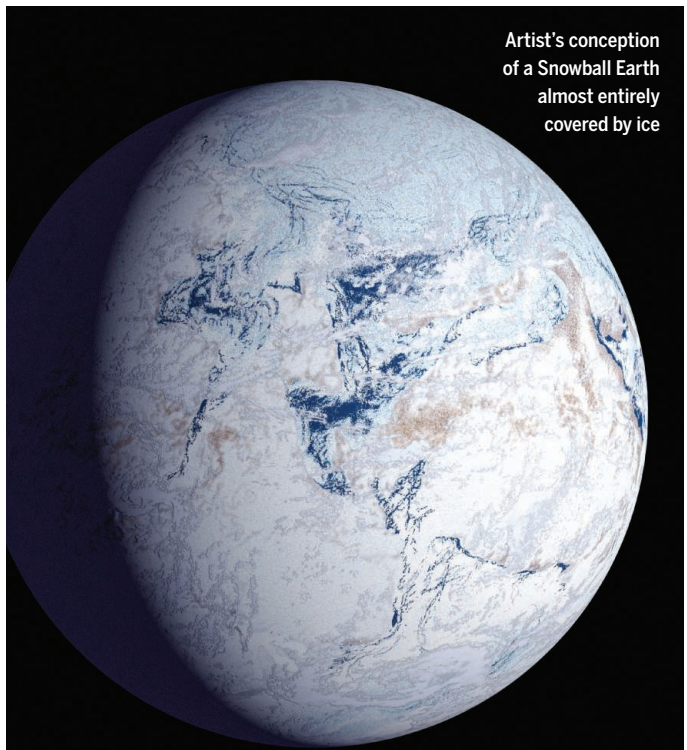
Sci. Transl. Med. **12**, eaaw3703 (2020).

PHOTO (FROM LEFT): SURADECH KONGKIATPAIBOON/ALAMY STOCK PHOTO; CHRIS BUTLER/SCIENCE SOURCE

Modern example of storage organs from a species of the *Hypoxis* genus, which includes the African potato

IN OTHER JOURNALS

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



Artist's conception of a Snowball Earth almost entirely covered by ice

SNOWBALL EARTH

Beating the freeze

Despite multiple episodes during the Cryogenian period in which the planet was mostly covered in ice for millions of years—essentially shutting down air–sea exchange and causing widespread ocean anoxia—marine aerobic eukaryotes somehow survived those Snowball Earth conditions. How did they do it? Lechte *et al.* show that oxygenated glacial meltwater was supplied to the oceans near ice sheet grounding lines, offering a lifeline to microbial communities there. This process may have created glaciomarine oxygen oases that were critical to the survival and evolution of those organisms during episodes of extreme glaciation. —HJS

Proc. Natl. Acad. Sci. U.S.A. **116**, 25478 (2019).

T CELL SIGNALING

Checkpoint cross-talk

Immunotherapy using antibodies that block the programmed cell death (ligand) 1 [PD-(L)1] or cytotoxic T lymphocyte–associated protein 4 (CTLA-4) immune checkpoint pathways has resulted in impressive responses for some cancer patients. Combined inhibition of both pathways has generally delivered better responses than the targeting of either alone, but how they functionally cross-talk is not well defined. Zhao *et al.* report that PD-L1 (the main ligand of PD-1) and CD80 (shared ligand for CTLA-4 and the central costimulatory receptor CD28), heterodimerize in cis. Using their model system, the authors found the PD-L1:CD80 cis complexes to be defective in binding either PD-1 or CTLA-4, but the ability of CD80 to activate CD28 appeared to be fully preserved.

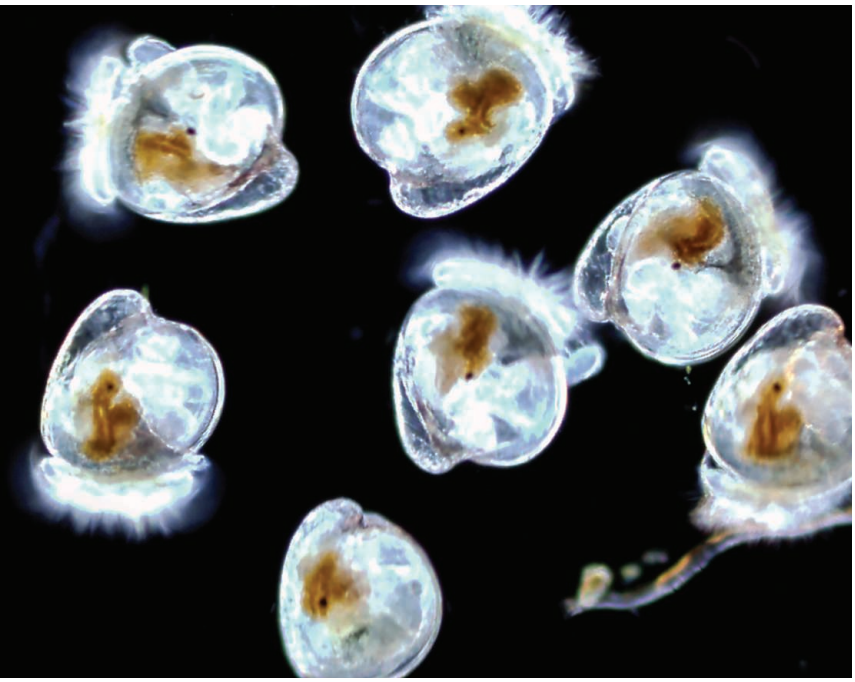
Published by AAAS

METALS IN BIOLOGY

Copper on the half shell

Oysters, like many marine mollusks, have a complex life cycle that requires changes in their body and biochemistry. Weng *et al.* examined how the abundance and distribution of copper, a trace element essential for metabolism, changes in larval and juvenile oysters as they develop. Using a combination of imaging techniques, they see copper accumulation in certain organs during developmental transitions, especially in the cells' mitochondria. Both cell proliferation and programmed cell death involve biochemical processes linked to mitochondrial copper concentrations. Understanding metal ion distributions may offer insights valuable for conservation and cultivation. —MAF

Environ. Sci. Technol. **53**, 14724 (2019).



Oysters have a particularly critical need for the trace element copper during their late pelagic larval stage (pictured).

PHOTO: STEPHANIE ALEXANDER, UNIVERSITY OF MARYLAND CENTER FOR ENVIRONMENTAL SCIENCE HORN POINT OYSTER HATCHERY

The interaction of checkpoint pathways might allow opportunities to improve therapy. —PNK

Immunity **51**, 1059 (2019).

NEUROSCIENCE

Filling a blank canvas

Natural scenes can be decoded in the brain in the early visual cortex. Decoding is a product of planned output (feedforward) corrected by experience (feedback). Morgan *et al.* occluded one quadrant of the visual field of human participants and then presented them with images of landscape scenes. Responses in the visual cortex (areas V1 and V2) were measured for each person by functional magnetic resonance imaging. Even in the occluded region, information about the scene was being decoded. To access what the brain had modeled for the occluded region, participants were asked to complete the image by line drawings on an electronic drawing tablet. The drawings were shown to all participants, who scored how well the "averaged" drawing for each scene matched the complete image. Mental models at the first visual cortical stage were predicted by the line drawings. So, faced with a blank canvas, an artist may be using cortical feedback processing until the lines and edges converge on her mental model of a scene, thus providing a window into her visual cortex. —PRS

J. Neurosci. **39**, 9410 (2019).

CELL BIOLOGY

Kill, but don't be killed

Cytotoxic T lymphocytes (CTLs) recognize and destroy infected or malignant cells. CTLs secrete the pore-forming protein perforin into the immune synapse, where perforin disrupts the target cell plasma membrane and initiates cell death pathways. Paradoxically, the secreted perforin only damages the target cell, leaving the producer lymphocytes unscathed and able to attack another target cell. What protects the lymphocyte is unclear. Using a variety of approaches, Rudd-Schmidt *et al.* found that two distinct but coordinated mechanisms were deployed by CTLs to protect themselves. Both mechanisms depended on the dynamic regulation of plasma membrane lipid composition and topology. High membrane lipid order made the CTL presynaptic membrane refractory to perforin binding. Furthermore, phosphatidylserine, exposed on the killer cell membrane, appeared to bind and inactivate perforin. —SMH

Nat. Commun. **10**, 5396 (2019).

MICROBIOTA

Postantibiotic recovery

Antibiotic consumption by humans is ubiquitous. Although they are lifesaving and valuable medicines, antibiotics can harm resident mutualistic microorganisms in unpredictable ways leading to physiological damage. Ng *et al.* systematically explored the long-term effects of antibiotic treatment on transplanted human microbiota in mice. The alpha-diversity of *Bacteroides* spp. suffered mass extinction by streptomycin within hours of treatment, whereas Firmicutes were more resilient. Such taxa-specific selection depended on the antibiotic used, influenced the recovered community composition, and affected resilience to subsequent treatments. Diet was an important variable: Without microbiota-accessible fiber in the diet, mucus grazers predominated, which may facilitate the activity of pathogens, such as *Acinetobacter muciniphila*. Recolonization from untreated cage mates was also vital for healthy reestablishment of the microbiota, especially of Bacteroidetes. It is possible that selective transplantation of extinct taxa could aid a healthy microbiota to recover after antibiotics exposure. —CA

Cell Host Microbe **26**, 650 (2019).

PHYSICS

Tailoring electronic properties

The layered material α-$RuCl_3$ has piqued the interest of physicists as a potential platform for the so-called Kitaev model—a type of quantum spin liquid that hosts exotic excitations. However, in addition to the magnetic interactions of the Kitaev model, α-$RuCl_3$ also hosts non-Kitaev interactions. Biswas *et al.* investigated theoretically whether a heterostructure consisting of a α-$RuCl_3$ monolayer on top of graphene would be a better approximation of the Kitaev model. Using ab initio calculations, the researchers found that the α-$RuCl_3$ layer would experience strain and become doped by graphene electrons. The strain would make the Kitaev interactions more dominant, and the doping may provide a route towards p-wave superconductivity in this system. —JS

Phys. Rev. Lett. **123**, 237201 (2019).

Published by AAAS

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

NEUROSCIENCE

The neural substrate of memory

The ability to form memory is an essential trait that allows learning and the accumulation of knowledge. But what is a memory? There has been a long history of searching for the neuronal substrate that forms memory in the brain, and the emerging view is that ensembles of engram cells explain how memories are formed and retrieved. In a Review, Josselyn and Tonegawa discuss the evidence for engram cells as a substrate of memory, particularly in rodents; what we have learned so far about the features of memory, including memory formation, retrieval over time, and loss; and future directions to understand how memory becomes knowledge. —GKA

Science, this issue p. 39

GLOBAL HEALTH

Cancer in diverse populations

Much of what is known about the signature genetic alterations and clinical features of cancer comes from studying Caucasian patients in high-income countries. However, this does not reflect the diversity of heritage in most societies and therefore does not allow patients to be diagnosed and treated effectively. In a Perspective, Rebbeck discusses the emerging evidence that cancers in patients in sub-Saharan Africa are clinically and molecularly distinct. Understanding these differences should improve treatment of patients in sub-Saharan Africa as well as the diaspora. Assessment of the diversity of genetic alterations can also improve how cancer is diagnosed and classified in all populations. —GKA

Science, this issue p. 27

MALARIA

An artemisinin resistance mechanism

Species of the malaria parasite *Plasmodium* live in red blood cells and possess a highly conserved gene called *kelch13*. Single point mutations in this gene are associated with resistance to the frontline artemisinin drugs. Birnbaum *et al*. found that Kelch13 and associated proteins comprise an endocytic compartment associated with feeding on host erythrocytes (see the Perspective by Marapana and Cowman). Hot targets for artemisinin research also occur in this compartment, including the proteins UBP1, AP-2μ, and the parasite homolog of the endocytosis protein Eps15. Inactivation of Kelch13 compartment proteins revealed that these are required for endocytosis of host hemoglobin. Artemisinins are activated by hemoglobin degradation products, so these mutations render the parasite resistant to these drugs to different extents. —CA

Science, this issue p. 51; see also p. 22

TOPOLOGICAL OPTICS

Optically contorting into new dimensions

Creating synthetic dimensions has generated interest in many branches of science, ranging from ultracold atomic physics to photonics. The ability to do so provides a versatile platform for realizing effective gauge potentials and novel topological physics that might be difficult or impossible to realize in real systems. Dutt *et al*. show that a structured optical ring cavity can sustain more than one synthetic dimension. Under modulation, coupling the different degrees of freedom within the resonator is used synthesize two additional dimensions. The authors are then able to emulate many complex physical phenomena usually associated with condensed matter systems. —ISO

Science, this issue p. 59

TOPOLOGICAL MATTER

Looking for chiral Majoranas

Chiral Majorana modes have been predicted to exist in heterostructures consisting of a quantum anomalous Hall insulator and a superconductor. Kayyalha *et al*. fabricated more than 30 such samples and used transport measurements to look for signatures of the Majorana modes. The data indicated that the transport signatures previously thought to be associated with Majorana physics could, in their samples, be explained using a more mundane mechanism. —JS

Science, this issue p. 64

TOPOLOGICAL MATTER

A possible propagating Majorana

Majorana states in solid-state systems may one day form a basis for topological quantum computing. Most of the candidates identified so far have been Majorana bound states, but theorists have predicted that propagating Majorana states may exist as well. Wang *et al*. looked for such a state on the surface of the material $FeSe_{0.45}Te_{0.55}$ (see the Perspective by Tewari and Stanescu). Using scanning tunneling spectroscopy, the researchers measured a flat, bias-independent density of states along a particular type of domain wall, which was consistent with a theoretical prediction for a propagating Majorana state in this material. Although topologically trivial origins of this finding are difficult to completely rule out, the work is likely to stimulate interest in iron-based superconductors as hosts of Majorana states. —JS

Science, this issue p. 104; see also p. 23

PLANT SCIENCE

Subtle origin for complex shapes

The cup-shaped leaves of carnivorous plants have evolved multiple times from ancestors with flat leaves. Studying development of the carnivorous trap in the humped bladderwort, *Utricularia gibba*, Whitewoods *et al*. identified genes similar to those expressed in surfaces of flat leaves (see the Perspective by Moulton and Goriely). Ectopic expression and computational modeling reveals how slight shifts in gene expression domains make the difference between a flat leaf and a convoluted trap structure. Flexibility in growth rates in orthogonal polarity fields allows for diversity in shapes formed through development. —PJH

Science, this issue p. 91; see also p. 24

TUBULIN

Mechanism of tubulin autoregulation

Cells tightly control the abundance of key housekeeping factors, such as ribosomes and chaperones, to maintain them at optimal levels needed for homeostasis. Most abundance control mechanisms involve feedback regulation of mRNA transcription, but others, such as tubulins, are regulated by highly specific mRNA degradation. Lin *et al.* found that tetratricopeptide protein 5 (TTC5) binds to nascent alpha and beta tubulins on translating ribosomes to trigger degradation of their associated mRNAs when excess tubulin is present (see the Perspective by Shoshani and Cleveland). In the absence of TTC5-mediated tubulin autoregulation, cells display error-prone chromosome segregation, a process critically dependent on tubulin concentration. —SMH

Science, this issue p. 100; see also p. 29

Published by AAAS

FLUORESCENT PROTEINS

Electrostatics guide chromophore twist

Photoisomerization—the twisting of bonds in a molecule in response to absorption of light—is exploited in biology to sense light and can influence the photophysical properties of fluorescent proteins used in imaging applications. Romei *et al*. studied this behavior by introducing unnatural amino acids into the photoswitchable green fluorescent protein Dronpa2, thus systematically altering the electronic properties of the chromophore (see the Perspective by Hu *et al*.). Crystal structures and spectroscopic analyses of a series of these variants support a model in which the electrostatic interactions between the chromophore and its environment influence the barrier heights for twisting around different bonds during photoisomerization. These insights may guide future design of photoswitchable proteins with desired properties. —MAF

Science, this issue p. 76; see also p. 26

CANCER

A movement orchestrated by DNA damage

Doxorubicin is a commonly used therapeutic that kills cancer cells by inducing genotoxic stress. Harvey *et al.* found that treatment of cancer cells with clinically relevant doses of doxorubicin promoted the phosphorylation of the translation factor eIF2α, which, rather than inhibiting protein synthesis as might be expected, led to increased migration. Inhibiting eIF2α phosphorylation with a U.S. Food and Drug Administration–approved drug restricted doxorubicin-induced cell migration. —WW

Sci. Signal. **12**, eaaw6763 (2019).

SUPERCONDUCTIVITY

Manipulating the gap

Cuprate superconductors typically have a large amount of spatial inhomogeneity, partly stemming from the inhomogeneity of the chemical doping process. In particular, the size of the spectroscopic gap can vary widely across a single sample. Massee *et al*. used the tip of a scanning tunneling microscope to manipulate atoms on the surface of a member of the $Bi_2Sr_2CaCu_2O_{8+x}$ cuprate family. Moving bismuth atoms up or down caused adjacent atoms to shift laterally, leading to reversible local changes in the size of the gap. It is expected that the technique can be used to probe the influence of the local lattice on the electronic states of other correlated materials. —JS

Science, this issue p. 68

Published by AAAS

REVIEW SUMMARY

NEUROSCIENCE

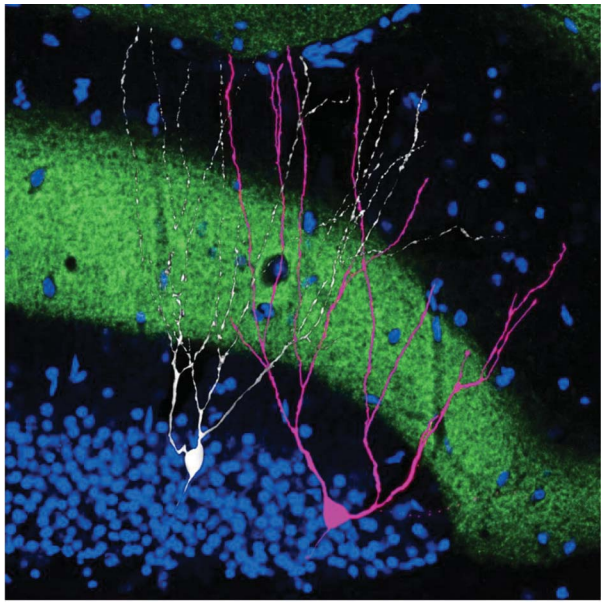
Memory engrams: Recalling the past and imagining the future

Sheena A. Josselyn* and Susumu Tonegawa*

BACKGROUND: The idea that memory is stored as enduring changes in the brain dates back at least to the time of Plato and Aristotle (circa 350 BCE), but its scientific articulation emerged in the 20th century when Richard Semon introduced the term "engram" to describe the neural substrate for storing and recalling memories. Essentially, Semon proposed that an experience activates a population of neurons that undergo persistent chemical and/or physical changes to become an engram. Subsequent reactivation of the engram by cues available at the time of the experience induces memory retrieval. After Karl Lashley failed to find the engram in a rat brain, studies attempting to localize an engram were largely abandoned. Spurred by Donald O. Hebb's theory that augmented synaptic strength and neuronal connectivity are critical for memory formation, many researchers showed that enhanced synaptic strength was correlated with memory. Nonetheless, the causal relationship between these enduring changes in synaptic connectivity with a specific, behaviorally identifiable memory at the level of the cell ensemble (an engram) awaited further advances in experimental technologies.

ADVANCES: The resurgence in research examining engrams may be linked to two complementary studies that applied intervention strategies to target individual neurons in an engram supporting a specific memory in mice. One study showed that ablating the subset of lateral amygdala neurons allocated to a putative engram disrupted subsequent memory retrieval (loss of function). The second study showed that artificially reactivating a subset of hippocampal dentate gyrus neurons that were active during a fearful experience (and, therefore, part of a putative engram) induced memory retrieval in the absence of external retrieval cues (gain of function). Subsequent findings from many labs used similar strategies to identify engrams in other brain regions supporting different types of memory.

There are several recent advances in engram research. First, eligible neurons within a given brain region were shown to compete for allocation to an engram, and relative neuronal excitability determines the outcome of this competition. Excitability-based competition also guides the organization of multiple engrams in the brain and determines how these engrams interact. Second, research examining the nature of the off-line, enduring changes in engram cells (neurons that are critical components of an engram) found increased synaptic strength and spine density in these neurons as well as preferential connectivity to other downstream engram cells. Therefore, both increased intrinsic excitability and synaptic plasticity work hand in hand to form engrams, and these mechanisms are also implicated in memory consolidation and retrieval processes. Third, it is now possible to artificially manipulate memory encoding and retrieval processes to generate false memories, or even create a memory in mice without any natural sensory experience (implantation of a memory for an experience that did not occur). Fourth, "silent" engrams were discovered in amnesic mice; artificial reactivation of silent engrams induces memory retrieval, whereas natural cues cannot. Endogenous engram silencing may contribute to the change in memory over time (e.g., systems memory consolidation) or in different circumstances (e.g., fear memory extinction). These findings suggest that once formed, an engram may exist in different states (from silent to active) on the basis of their retrievability. Although initial engram studies focused on single brain regions, an emerging concept is that a given memory is supported by an engram complex, composed of functionally connected engram cell ensembles dispersed across multiple brain regions, with each ensemble supporting a component of the overall memory.



An engram cell alongside a nonengram cell. Within the hippocampus, dentate gyrus cells were filled with biocytin (white) to examine morphology. Engram cells active during context fear conditioning were engineered to express the red fluorescent protein mCherry, which appears pink owing to overlap with biocytin signals. Axons of the perforant path (green) express the excitatory opsin channelrhodopsin 2 and a fluorescent marker (enhanced yellow fluorescent protein). The upper blade of the dentate gyrus granule cell layer is revealed by the nuclear stain 4′,6-diamidino-2-phenylindole (DAPI, blue).

OUTLOOK: The ability to identify and manipulate engram cells and brain-wide engram complexes has introduced an exciting new era of memory research. The findings from many labs are beginning to define an engram as the basic unit of memory. However, many questions remain. In the short term, it is critical to characterize how information is stored in an engram, including how engram architecture affects memory quality, strength, and precision; how multiple engrams interact; how engrams change over time; and the role of engram silencing in these processes. The long-term goal of engram research is to leverage the fundamental findings from rodent engram studies to understand how information is acquired, stored, and used in humans and facilitate the treatment of human memory, or other information-processing, disorders. The development of low- to noninvasive technology may enable new human therapies based on the growing knowledge of engrams in rodents. ■

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aaw4325

The list of author affiliations is available in the full article online.
*Corresponding author. Email: sheena.josselyn@sickkids.ca (S.A.J.); tonegawa@mit.edu (S.T.)
Cite this article as S. A. Josselyn and S. Tonegawa, *Science* **367**, eaaw4325 (2020). DOI: 10.1126/science.aaw4325

CREDIT: ADAPTED FROM T. J. RYAN *ET AL.*, *SCIENCE* **348**, 1007 (2015).

REVIEW

NEUROSCIENCE

Memory engrams: Recalling the past and imagining the future

Sheena A. Josselyn[1,2,3,4,5]* and Susumu Tonegawa[6,7]*

In 1904, Richard Semon introduced the term "engram" to describe the neural substrate for storing memories. An experience, Semon proposed, activates a subset of cells that undergo off-line, persistent chemical and/or physical changes to become an engram. Subsequent reactivation of this engram induces memory retrieval. Although Semon's contributions were largely ignored in his lifetime, new technologies that allow researchers to image and manipulate the brain at the level of individual neurons has reinvigorated engram research. We review recent progress in studying engrams, including an evaluation of evidence for the existence of engrams, the importance of intrinsic excitability and synaptic plasticity in engrams, and the lifetime of an engram. Together, these findings are beginning to define an engram as the basic unit of memory.

Memory is the ability to use the past in service of the present or future (*1*, *2*). Memory is central to our everyday lives and defines who we are. Without it, we are condemned to an eternal present. That memory persists after an experience suggests that an internal representation of this experience is stored in the brain and that later this representation can be reconstructed and used. In 1904, Richard Semon, an evolutionary zoologist turned memory theorist, introduced the term "engram" to describe such memory representations (*3*, *4*). Semon defined an engram as "…the enduring though primarily latent modifications in the irritable substance produced by a stimulus…" (*5*, p. 12; *6*). He postulated a fundamental "law of engraphy" in which "all simultaneous excitations… form a connected simultaneous complex of excitations which, as such, act engraphically, that is to say leaves behind it a connected, and to that extent, unified engram-complex" (*7*, p. 159–160). An engram, therefore, is roughly equivalent to a "memory trace."

Semon's innovative ideas were largely overlooked or dismissed during his lifetime. However, his theories foreshadowed many prominent contemporary memory concepts (*8*–*11*). Semon defined an engram as an off-line, physical change in some aspect of brain state but was suitably cautious when asked to speculate on the precise neural mechanisms underlying an engram, "To follow this into the molecular field seems to me…a hopeless undertaking at the present stage of our knowledge and for my part, I renounce the task" (*7*, p. 154).

A few years later, though, Karl Lashley, a geneticist turned psychologist, took up this challenge by systematically attempting to localize an engram in a mammalian brain (*12*–*14*). In a typical study, Lashley trained rats over many days to solve a maze by running a distinct route to collect a reward. Hypothesizing that some critical component of the engram supporting this maze-route memory is localized in the cortex, Lashley removed cortical tissue of varying sizes from varying locations and then tested the rats' memory for the maze route. Although the amount of cortical tissue removed correlated with overall memory impairment, the location of the lesion did not. After more than 30 years of searching, Lashley failed to find an engram, declaring it "elusive."

The next leap in engram-related research came when Donald O. Hebb, a psychologist, memory theorist, and student of Lashley, developed a cell assembly theory (similar to Semon's engram complex) (*15*). Hebb hypothesized that a cell assembly is formed between reciprocally interconnected cells that are simultaneously active during an experience. Sufficient activity within the cell assembly induces growth and/or metabolic changes that strengthen the connections between these cells [a concept distilled in the phrase "neurons that fire together, wire together" (*16*)]. These synaptic and metabolic changes (perhaps including changes in intrinsic neuronal excitability) have implications for the function of a cell assembly. For instance, reactivation of only a fraction of assembly cells was hypothesized to produce reactivation of the entire assembly (*15*) [a process similar to pattern completion (*17*–*19*)]. By contrast, destruction of a fraction of assembly cells would not necessarily produce catastrophic failure of the entire representation (but rather gracefully degrade the representation). Interestingly, Semon also proposed similar types of properties for an engram (*5*).

Together, these (and other) scientists helped define and describe an engram. However, there was a paucity of studies examining the biological basis of engrams. More than 100 years ago, Semon wrote that to examine the neurobiological basis of an engram represented a "hopeless undertaking." This may no longer be true. Recent excitement surrounding engram research may stem directly from the development of new tools allowing cell ensembles to be imaged and manipulated at the level of the individual cell. We begin by briefly reviewing the neurobiological evidence supporting the existence of engrams in the rodent brain and our collective ability to not only find but also manipulate engrams to better understand memory. Then, we discuss the current state of engram research by examining the results of explicit engram studies and previous memory and plasticity findings from an engram point of view. Guided by Semon, we define an engram as an enduring off-line representation of a past experience (Box 1). It is important to note that an engram is not yet a memory but rather provides the necessary physical conditions for a

Box 1. Engram definitions.

An "engram" refers to the enduring offline physical and/or chemical changes that were elicited by learning and underlie the newly formed memory associations.

"Engram cells" are populations of cells that constitute critical cellular components of a given engram. These cells may (or may not) also be critical components of engrams supporting other memories. Engram cells are (i) activated by a learning experience, (ii) physically or chemically modified by the learning experience, and (iii) reactivated by subsequent presentation of the stimuli present at the learning experience (or some portion thereof), resulting in memory retrieval.

An "engram cell ensemble" refers to the collection of engram cells localized within a brain region. Engram cell ensembles in each brain region are connected, forming an "engram complex," which is the entire brainwide engram supporting a memory that is stored in sets of engram cell ensembles in different brain regions connected via an engram cell pathway.

[1]Program in Neurosciences & Mental Health, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. [2]Department of Psychology, University of Toronto, Toronto, Ontario M5S 3G3, Canada. [3]Department of Physiology, University of Toronto, Toronto, Ontario M5G 1X8, Canada. [4]Institute of Medical Sciences, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [5]Brain, Mind & Consciousness Program, Canadian Institute for Advanced Research (CIFAR), Toronto, Ontario M5G 1M1, Canada. [6]RIKEN-MIT Laboratory for Neural Circuit Genetics at the Picower Institute for Learning and Memory, Department of Biology and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [7]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
*Corresponding author. Email: sheena.josselyn@sickkids.ca (S.A.J.); tonegawa@mit.edu (S.T.)

memory to emerge (20). Memories are retrieved when appropriate retrieval cues successfully reactivate an engram in a process Semon dubbed "ecphory."

Experimental strategies to evaluate engrams

To evaluate the existence of engrams, we adapt the criteria and experimental strategies discussed by Morris and colleagues (21, 22) in their landmark papers evaluating the importance of synaptic plasticity in memory. Specifically, we discuss evidence from four types of studies. First, observational studies supporting the existence of engrams in the rodent brain should show that the same (or overlapping) cell populations are activated both by an experience and by retrieval of that experience and that, furthermore, learning should induce long-lasting cellular and/or synaptic modifications in these cells. Second, loss-of-function studies should show that impairing engram cell function after an experience impairs subsequent memory retrieval. Third, gain-of-function studies should show that artificially activating engram cells induces memory retrieval, in the absence of any natural sensory retrieval cues. Fourth, mimicry studies should artificially introduce an engram of an experience that never happened into the brain and show that rodents use the information of an artificial engram to guide behavior.

Memory traces, or at least physiological correlates of memory, have been examined in invertebrate species, such as flies (23–27), octopus (28, 29), *Aplysia* sea slugs (30, 31), honey bee (32), and *Hermissenda* sea slugs (33). Moreover, pioneering studies in mammals (34–36) greatly informed our current understanding of the neural basis of memory but did not examine memory at the cell ensemble level. The discussion here is limited primarily to rodent experiments examining memory of an explicit experience that probe memory at the level of an engram.

Observational studies

Typically, observational studies take advantage of immediate early genes (IEGs) such as *c-Fos*, *Arc* (activity-regulated cytoskeleton-associated protein), or *Zif268* (zinc finger protein 225) (37–39) to visualize active neurons. Cells active during a memory test are marked using IEG immunohistochemistry, whereas cells active during a training experience are "tagged" through the use of temporally inducible IEG promoters that drive the expression of more enduring fluorescent (or other) reporter proteins (40–43). Above-chance overlap between these two cell populations ("active during training" and "active during test") within a brain region (or throughout the brain) is suggestive of an engram.

In an initial observational study designed to examine a memory at the level of a cell ensemble, Mayford and colleagues (41) tagged neurons active during auditory fear conditioning. In this commonly used memory task, an initially innocuous tone (a conditioned stimulus) is paired with an aversive footshock (an unconditioned stimulus) in a conditioning context. When subsequently reexposed to the tone or conditioning context, rodents freeze (the active, learned conditioned response), showing memory of the training experience (44). In this experiment, mice were replaced in the conditioning context 3 days after training, and active neurons were marked with zif268 immunohistochemistry. Consistent with the existence of an engram supporting this conditioned fear memory, the overlap of neurons active during training (tagged) and testing (zif268+) in the basal amygdala nucleus exceeded chance (~11% total cells) (41).

Similar results, using different tagging methods, across multiple brain regions [including dorsal hippocampus (40, 45–55), amygdala (41, 45, 49, 51, 55, 56), and cortex (42, 45, 55, 57)] were reported for a variety of different memory tasks (including contextual fear conditioning, auditory fear conditioning, and novel object exploration). Control studies revealed that tagged cells were only reactivated by the corresponding conditioned stimulus and not by stimuli unrelated to the training experience (45). Although most observational studies did not address directly the enduring, learning-induced changes hypothesized by Semon, overall, these results (and their notable consistency across methods, tasks, and labs) provide broad support for the existence of engrams. However, causal studies are necessary to show that these reactivated putative engram cells indeed function as part of the internal representation of an experience.

Loss-of-function studies

Loss-of-function studies attempt to "capture" engram cells and specifically disrupt their function before a memory test. Josselyn and colleagues (58) performed the first loss-of-function memory study at the level of a cell ensemble. An allocation strategy was used to capture putative engram cells in the amygdala lateral nucleus (LA) supporting an auditory fear conditioned memory in mice. That is, a small, random population of LA neurons was biased for inclusion (or allocation) into a putative engram using a neurotropic virus expressing CREB (Ca^{++}/cyclic AMP–responsive element-binding protein). CREB is a transcription factor that increases both neuronal excitability (59–64) and dendritic spine density (60, 65). Therefore, neurons infected with this CREB vector were hypothesized to be biased for inclusion into an engram. A virus expressing both CREB (to allocate neurons) and an inducible construct that produces cell-autonomous ablation was used to specifically kill allocated neurons after training (58). Ablating CREB-overexpressing neurons disrupted freezing to subsequent tone presentation, as if the memory was erased (Fig. 1). Importantly, mice were capable of learning a new fear conditioning task (showing overall LA function was not compromised), and ablating a similar number of non–CREB-overexpressing cells (nonengram cells) did not disrupt memory (showing specificity of the memory disruption at the cellular level).

Subsequent studies using diverse methods to permanently or reversibly inactivate allocated or tagged neurons across several brain areas hypothesized to be part of an engram, in many memory tasks, produced comparable results (40, 48, 53, 63, 66, 67). Together, these findings suggest that neurons active during an experience become engram cells that are indispensable (or somehow necessary) for successful subsequent memory expression.

Why were these loss-of-function studies perhaps successful in "finding an engram" when Lashley was not? First, Lashley may have used an inappropriate behavioral test to probe an engram. The well-learned maze task Lashley typically used could be solved using different strategies and, therefore, may have been insensitive to damaging a distinct brain region. Second, Lashley may have targeted the wrong brain region for this type of spatial memory task (68).

Gain-of-function studies

Gain-of-function studies attempt to induce memory retrieval in the absence of natural retrieval cues by artificially reactivating engram cells. Tonegawa and colleagues (69) provided the first gain-of-function evidence for the existence of an engram. Hippocampal dentate gyrus (DG) neurons active during contextual fear conditioning (in which a context was paired with a footshock) were tagged (41) to express the excitatory opsin channelrhodopsin 2 (ChR2) (70). When tested in a nontraining context, mice did not freeze. However, photostimulation of tagged engram cells was sufficient to induce freezing, the learning-specific conditioned response (44), even though mice had never been shocked in this nontraining context (Fig. 2). Importantly, light-induced freezing was not due to activation of prewired learning-independent neural circuits or a simple reflex response, because similar photostimulation of tagged DG neurons failed to induce freezing if downstream CA1 neurons were silenced during training (thereby preventing learning) (71).

Artificial optogenetic or chemogenetic (72, 73) reactivation of tagged or allocated engram cells across several brain regions similarly induced memory expression without external sensory retrieval cues in a variety of tasks (42, 53, 74–81). Therefore, artificial engram cell reactivation

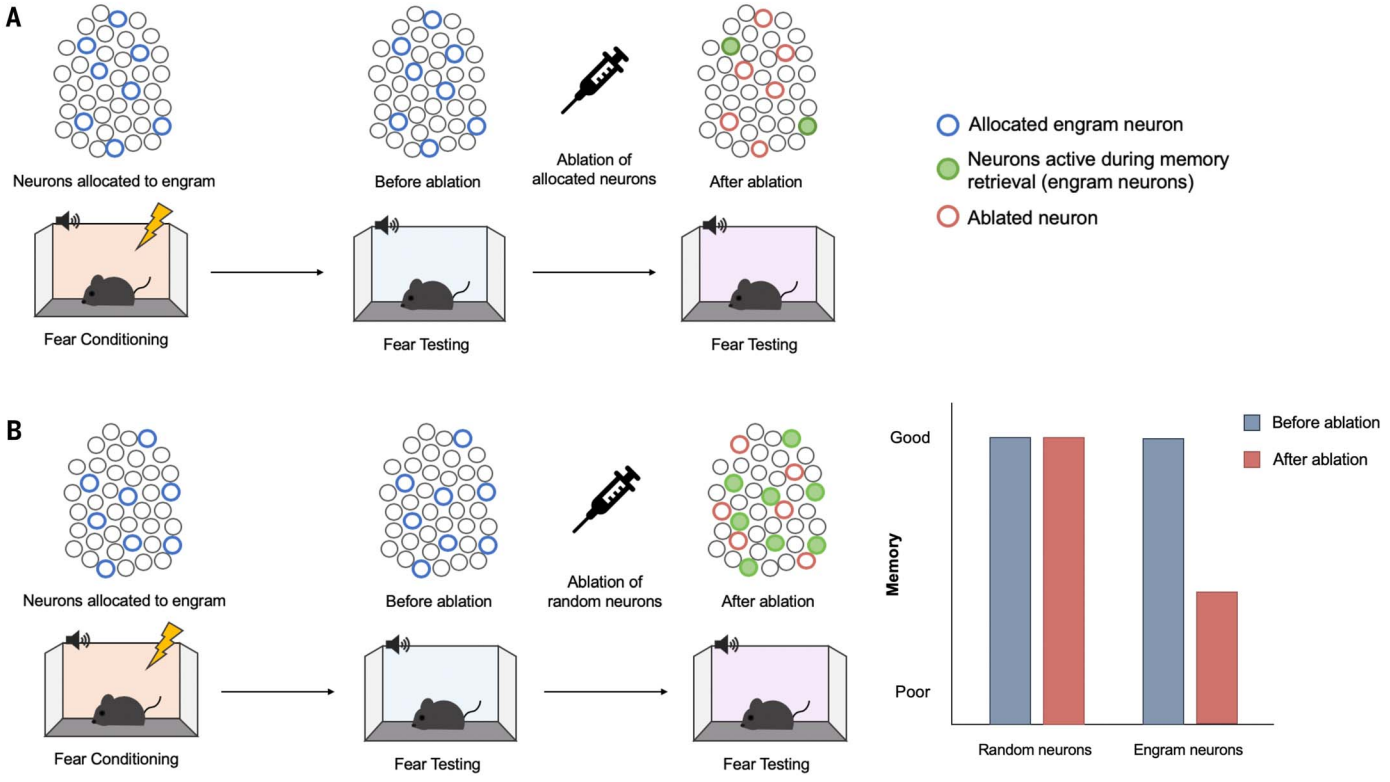


Fig. 1. Engram loss-of-function studies disrupt subsequent memory retrieval. (**A**) Ablating allocated neurons. Lateral amygdala principal (excitatory) neurons were experimentally allocated to an engram (blue circles) by means of overexpression of the transcription factor CREB (*122*). Mice received auditory fear conditioning during which a tone (conditioned stimulus) was paired with a footshock (unconditioned stimulus). The majority of allocated neurons are active during the fear memory test (green filled circles), suggesting that allocated neurons are preferentially recruited to an engram supporting this conditioned fear memory. Specifically ablating experimentally allocated neurons (red circles) before a second memory test disrupts memory retrieval. (**B**) Ablating a similar number of random, nonallocated neurons does not disrupt memory retrieval. [Images: Adapted from (*122*)].

serves as a sufficient retrieval cue to "reawaken" a dormant engram to induce memory expression, similar to Semon's original definition of ecphory ["the influences which awaken the mnemic trace or engram out of its latent state into one of manifested activity" (*5*, p. 12)].

Mimicry experiments

During natural memory retrieval, the sensory conditioned stimulus (e.g., the training context) is thought to reactivate engram cells to induce memory retrieval. The first gain-of-function study (*69*) was designed to mimic this retrieval process by directly reactivating engram cells by means of optogenetic stimulation, thus circumventing the need for the conditioned stimulus. That is, artificial stimulation replaced the natural conditioned stimulus to induce memory retrieval. Optogenetic stimulation of engram cells has also been used to artificially retrieve a previously experienced sensory stimulus during the formation of a new memory. For instance, DG neurons active during exploration of a new context (context A) were photostimulated when mice later received footshocks in a different context (context B). During a memory test, mice replaced in context A froze, even though they had never been shocked in this context. That is, mice retrieved an artificial memory. Mice also froze in context B (showing natural memory retrieval), but not in a third distinct context (context C), indicating freezing was a context-specific, and not a generalized, response (*46*). Both memories produced by "natural" and "artificial" means could only be retrieved by their respective conditioned stimuli, indicating both memories retained their identities. Similar to a compound conditioned stimulus in which both a tone and light predict footshock, the strength of the natural and artificial memories were roughly 50% of a single "normally induced" memory, suggesting cue competition between the natural and artificial conditioned stimuli [as originally described by (*82*)]. Therefore, when a biologically important event (e.g., footshock) occurs while an animal is retrieving a previously formed but perhaps unrelated memory, the two stimuli can be associated to form a new but false episodic memory. An analogous mechanism may underlie human false memories, except that in humans, the previously acquired memory would be retrieved by natural processes (*83*).

Mayford and colleagues (*84*) used a similar approach but tagged active neurons across the brain as mice explored a new context (context A). Chemogenetically reactivating these neurons while mice were fear conditioned in context B produced a "hybrid or synthetic" context representation that was not retrievable by either context alone [unlike (*46*), above]. However, mice froze in a test session that more closely matched the training conditions (placement in context B while chemogenetically activating context A engram cells), suggesting that this hybrid memory incorporated both natural and artificial cues. Differences in the spatial and temporal properties of artificial engram reactivation (more acute optogenetic activation of localized tagged DG neurons versus longer-term chemogenetic activation of nonlocalized tagged neurons across the brain) may account for the discrepant outcomes of these two artificial conditioned stimuli studies.

Neurons active during presentation of an unconditioned stimulus have also been tagged and artificially reactivated (*85*). Neuronal ensembles active during context exploration (the conditioned stimulus) and footshock (the unconditioned stimulus) were tagged separately in the CA1 subfield of the hippocampus and the basolateral complex of the amygdala, respectively. Synchronous optogenetic activation

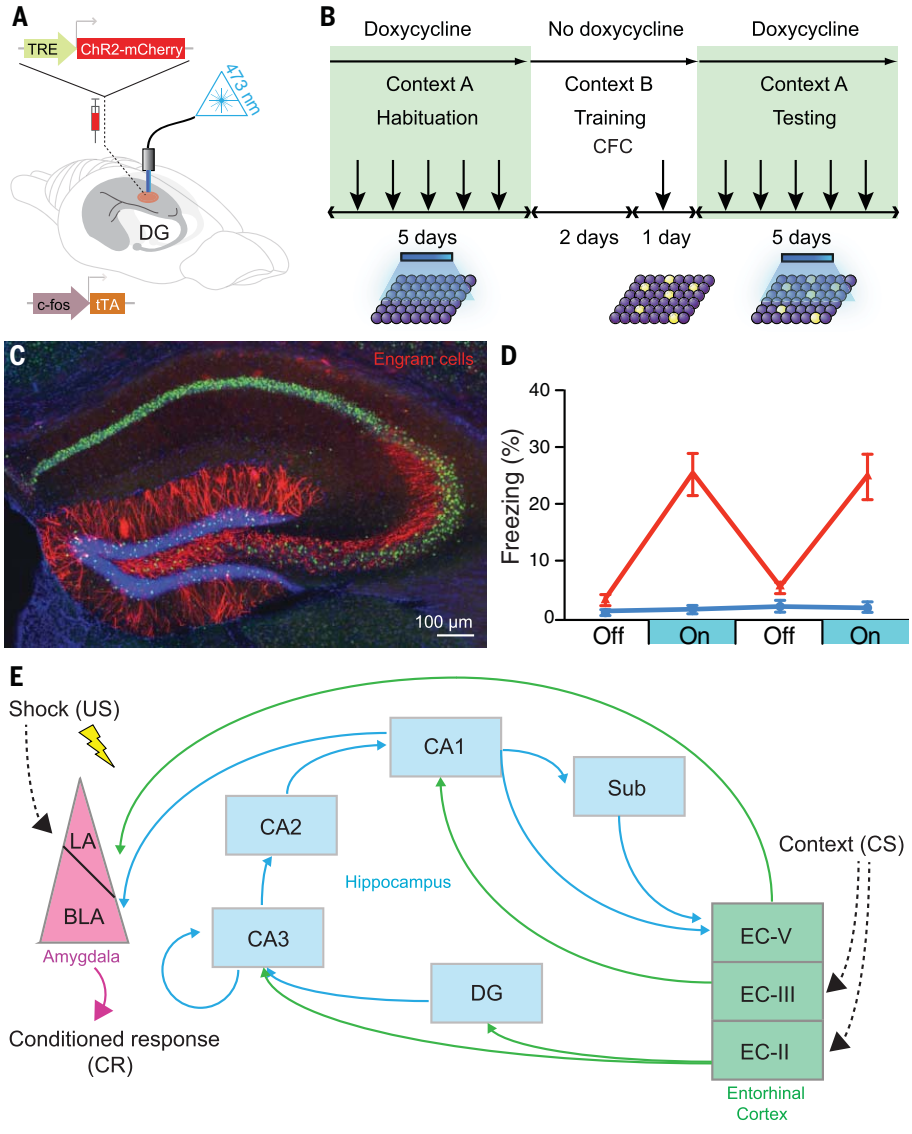


Fig. 2. Gain-of-function method for engram identification and distributed engram ensembles. (**A**) A c-fos–tTA transgenic mouse is injected with AAV9-TRE-ChR2-mCherry (allowing active neurons in the absence of doxycycline to express the excitatory opsin ChR2) and implanted with an optical fiber to target blue light to activate ChR2-expressing neurons in the DG. (**B**) Basic experimental scheme. Mice are habituated to context A with light stimulation while on doxycycline for 5 days and are then taken off doxycycline for 2 days (to open the tagging window) and exposed to contextual fear conditioning (CFC) in context B. Mice are put back on doxycycline (to close the tagging window) and tested for 5 days in context A with light stimulation. (**C**) Representative image showing the expression of ChR2-mCherry–positive (red) engram cells in a mouse that was taken off doxycycline for 2 days and underwent CFC training. [Image credit: X. Liu and S. Ramirez (Tonegawa lab)] (**D**) Mice expressing ChR2 in engram cells from CFC in context B (red) show greater freezing during test light-on epochs in context A than a control group expressing mCherry only. Error bars indicate standard error of the mean. [Graph: Adapted from Liu *et al.* (*69*)] (**E**) A part of the engram cell ensemble complex for contextual fear memory. It is generally thought that the engram for a specific memory is distributed in more than one brain region. For instance, for contextual fear memory, the engram cell ensemble in the entorhinal cortex layer II (EC-II) as well as hippocampal subfields [DG, CA3, CA2, CA1, and subiculum (Sub)] may represent context, whereas amygdala engram cell ensembles represent fear information. These engram cell ensembles are functionally connected to form an engram cell ensemble complex. Thus, a concept has emerged that a specific pattern of cellular connectivity within an engram cell ensemble complex serves as the substrate for a specific memory. US, unconditioned stimulus; LA, lateral nucleus of the amygdala; BLA, basolateral nucleus of the amygdala; CS, conditioned stimulus.

of these ensembles while mice were in the homecage was sufficient to induce a false memory; mice froze in the tagged (but nonshocked) context, as if the conditioned stimulus and unconditioned stimulus had been paired.

Finally, a recent study investigated whether a memory could be implanted through artificial means in the total absence of natural stimuli (either conditioned stimulus or unconditioned stimulus). To be a true memory implantation, such an experiment should satisfy several criteria (*86*). First, the "learning experience" should occur entirely within the brain through, for example, direct stimulation of putative conditioned-stimulus and unconditioned-stimulus neural pathways. Second, the presence of the implanted memory should be probed through presentation of a "real" external retrieval cue (not just the internal neural cue). Finally, behavioral manifestation of this memory should reflect the predicted memory content and be retrieved only by the "trained" conditioned stimulus (not to similar cues). In this study, optogenetic stimulation of a genetically specific olfactory glomerulus (the conditioned stimulus) was paired with optogenetic stimulation of either appetitive or aversive neural pathways (the unconditioned stimuli) (*86*). After this entirely intracranial conditioning, mice showed either an attraction or aversion, respectively, to the real odor that activated this olfactory glomerulus. In short, a memory was made in the absence of experience. These results satisfy the mimicry criterion of experimental evidence outlined by Martin and colleagues (*21*, *22*) and, as such, provide another line of persuasive evidence for the existence of engrams.

Understanding memory through engrams

The "enduring changes" of an engram

The ability to label in vivo engram cells supporting a specific memory provided an opportunity to investigate the nature of the "enduring changes" proposed by Semon. Guided by Hebb's influential theory on the critical importance of synaptic plasticity (the increase in synaptic strength between neurons) in memory [e.g., (*21*, *22*)], Tonegawa and colleagues showed that learning augmented synaptic strength, specifically in engram cells. First, 1 day after training, hippocampal DG granule engram cells tagged during contextual fear conditioning showed greater synaptic strength [higher AMPA/NMDA ratio, which is a means of assessing basal strength of excitatory synapses by examining the relative expression of amino-3-hydroxy-5-methyl-4-isoxazole propionic acid receptor (AMPAR)–mediated synaptic currents to *N*-methyl-D-aspartate receptor (NMDAR)–mediated synaptic currents of a population of stimulated synapses (*87*)] and increased spine density at entorhinal cortex junctions than nonengram DG cells (*71*).

Second, compared to nonengram CA3 cells, downstream CA3 engram cells were more functionally connected with upstream DG engram cells (*71*). Moreover, Kaang and colleagues showed that the number and sizes of spines on CA1 engram cells tagged during contextual fear conditioning receiving input from CA3 engram cells was greater than on nonengram CA1 cells. This enhanced interregional connectivity between CA3 and CA1 engram cells correlated with memory strength and occluded long-term potentiation (LTP), suggesting a previous LTP-like phenomenon endogenously occurred (*88*). Similarly, LA engram cells tagged during auditory fear conditioning showed enhanced synaptic connectivity with presynaptic neurons (*56*, *89*). Finally, shrinking potentiated synapses in primary motor cortex (M1) engram cells supporting a motor memory disrupted subsequent performance of this, and not a similar, motor memory (*90*). Together, these studies are beginning to integrate previous research on synaptic plasticity with engrams and suggest preferential engram cell–to–engram cell connectivity is a critical part of the enduring changes to an engram generated by learning. Overall these findings suggest an update of Hebb's axiom: Engram cells that fire together, wire together.

Distributed engram ensembles

Although one specific brain region is often examined in engram studies, it is generally appreciated that an engram supporting a specific experience may be widely distributed throughout the brain. Engram cell ensembles in different brain regions may support distinct aspects of an experience. For instance, in contextual fear memory, hippocampal (DG, CA3, and CA1) engram cell ensembles may represent the context (*40*, *48*, *91*–*93*), whereas amygdala engram cell ensembles may represent valence information (*69*, *71*, *75*), and cortical engram cell ensembles may represent distinct sensory information (*79*, *94*–*96*).

Several studies have examined potential engram cell ensembles supporting contextual fear memories across the brain (*42*, *97*–*99*). For instance, Frankland and colleagues compared the brainwide (84 brain regions) distribution of active cells after retrieval of recent (1 day after training) versus remote (36 days after training) contextual fear memory. On the basis of coactivation, graph theory was used to construct functional connectome "memory maps" (*97*) and identify hub-like regions hypothesized to play privileged roles in memory retrieval. Subsequent chemogenetic inhibition confirmed that these identified hub regions were necessary for subsequent memory retrieval (*98*). Using a combination of engram tagging technology [targeted recombination in active populations 2 (TRAP2) transgenic mice] and IEG immunohistochemistry to examine overlap between neurons active at contextual fear training and testing, Luo and colleagues (*42*) showed that retrieval of a remote (14 day) contextual fear memory engaged more neurons in prelimbic cortex than retrieval of a recent (1 day) memory, suggesting that an engram changes over time [consistent with the findings of (*100*)]. Finally, a preliminary study (*99*) mapped candidate engram ensembles representing a contextual fear conditioning memory in 409 brain regions in mice. Roy and colleagues tagged cells active at training and those active at recall throughout the brain in the same mouse using a CLARITY-like tissue-clearing technique (*101*) dubbed SHIELD (stabilization under harsh conditions via intramolecular epoxide linkages to prevent degradation) (*102*), thereby permitting the entire intact brain to be imaged at once. From this activation data, these researchers developed an "engram index" (defined as the degree to which cells in a given brain region were active at memory encoding and retrieval) that allowed the rank ordering of different brain regions. Using optogenetic and chemogenetic methods to interrogate the effects of artificially activating regions with a high engram index, this study showed many of these engram ensembles are functionally connected and activated simultaneously by an experience. These findings suggest that an experience is represented in specifically connected multiple engram ensembles distributed across multiple brain regions and provide experimental support for Semon's "unified engram complex" hypothesis.

Engrams, place cells, and sleep

Location-specific firing of CA1 place cells is well established (*103*). Stable place cells may be important in engrams supporting spatial or contextual memories (*104*–*106*). Recently, McHugh and colleagues (*107*) contrasted the roles of CA1 place cells and engram cells in memory. While mice explored a new context, engram cells were tagged and place cells identified using tetrode recordings. Most tagged engram cells were also place cells, but the majority of place cells were not tagged. Nontagged place cells behaved like traditional place cells (stable in the same context but remapping in a new context). By contrast, tagged place cells fired in a context-specific manner, albeit with imprecise spatial information, and were not active (did not remap) in a new context. Therefore, engram cells may provide general contextual information, with nontagged place cells providing precise spatial information.

Postencoding reactivation or replay of hippocampal place cell firing, especially during slow-wave sleep (SWS) (*108*, *109*), is thought to be important for memory consolidation (*110*–*113*). During SWS, hippocampal neurons fire in an oscillatory rhythm (termed sharp-wave ripples), tending to co-occur with rhythmic firing of cortical neurons (termed spindles) (*114*). Disrupting either sharp-wave ripple–spindle coupling (*115*, *116*) or sharp-wave ripple–associated replay of hippocampal place cells (*104*, *105*, *117*, *118*) impairs memory recall. The precise role of these rhythmic oscillations with respect to engram cells is unclear. Sharp-wave ripples promote synaptic depression of CA1 hippocampal neurons (*119*, *120*). A recent study suggests that CA1 engram cells tagged during context exploration are more likely than nonengram neurons to participate in sharp-wave ripple events, perhaps allowing these engram cells to escape this SWS-induced synaptic depression (*120*). In this way, postencoding reactivation of engram cells during oscillatory rhythms may help refine an engram by decreasing irrelevant "noise" of nonengram neuronal activity during memory consolidation.

Lifetime of an engram

Birth of an engram

Josselyn, Silva, and colleagues discovered that during engram formation, eligible neurons in a given brain region compete against each other for allocation (or recruitment) to an engram. Neurons with relatively increased intrinsic excitability win this allocation competition to become engram cells (*58*, *63*, *66*, *76*, *77*, *121*–*126*) (Fig. 3). Competitive excitability-based allocation to an engram occurs in other brain regions and supports different types of memories [e.g., dorsal CA1 region of hippocampus (*91*–*93*) and prefrontal cortex (*126*) (for a contextual fear memory), insular cortex (*127*) (conditioned taste-aversion memory), and retrosplenial cortex (*128*) (spatial memory)].

In addition to aversive memories, LA neurons experimentally made more excitable during training were also preferentially allocated to an engram supporting a cocaine-cue rewarding memory (*66*). Similarly, increasing the excitability of a small, random portion of piriform cortex principal neurons resulted in their allocation to an engram supporting either a rewarding or an aversive olfactory memory, depending on the nature of the training experience (*129*). Excitability-based neuronal allocation is predicted by computational modeling (*130*–*132*), occurs endogenously (*56*, *89*), and is consistent with previous research implicating intrinsic excitability in the formation of invertebrate memory traces (*33*, *133*–*135*). Together, these findings suggest that in some brain regions, at any given time, a small portion of eligible neurons are "primed" to become part of an engram (should an experience occur), regardless of experience valence.

Although stable place cells and engram cells in dorsal CA1 of the hippocampus differ (*107*), some mechanisms underlying their formation may be shared. In a given environment, only a small subset of CA1 neurons are place cells,

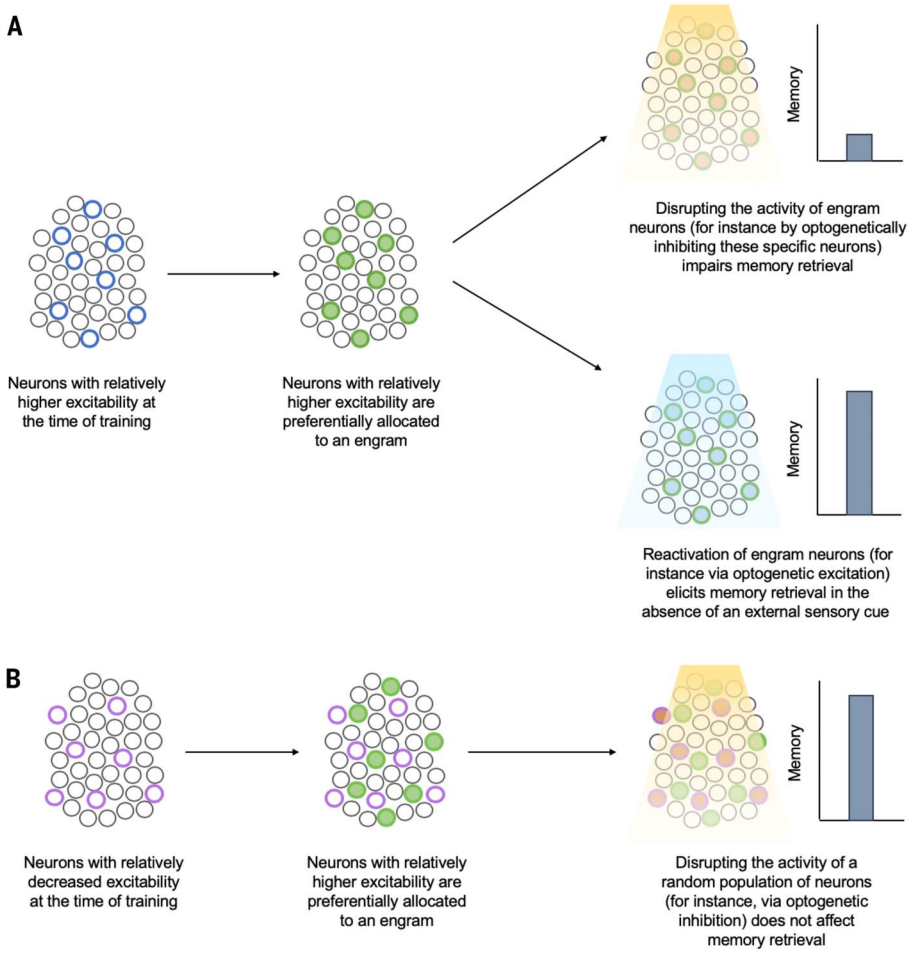


Fig. 3. Neuronal allocation to an engram. Eligible neurons compete for allocation to an engram supporting a memory, and neurons with increased relative excitability at the time of training "win" this competition for allocation. (**A**) Neurons that were endogenously more excitable than their neighbors at the time of training or were experimentally manipulated to become relatively more excitable (blue circles) are preferentially allocated to an engram (green filled circles). Subsequent disruption of these allocated or engram neurons disrupts memory retrieval (top right), whereas artificial reactivation of these neurons elicits memory retrieval in the absence of normal sensory retrieval cues (bottom right). (**B**) Neurons with relatively decreased excitability at the time of training (either endogenously or through experimental manipulation) (purple circles) are preferentially excluded from the engram (green filled circles). Subsequent disruption of nonallocated or nonengram neurons does not impact memory retrieval.

because the majority of CA1 neurons are silent (*136*). Those neurons with relatively higher excitability immediately before placement in a novel environment are more likely to become place cells in that environment (*137*–*139*), and experimentally increasing the excitability of an initially silent cell biased this cell toward becoming a place cell (*140*, *141*).

It is interesting to note the similarities between data from current allocation studies and the long-standing idea of selective stabilization (*142*, *143*). Selective stabilization proposes that multiple prerepresentations are endogenously generated in the brain and only one or a few that fit the situation are selected at any given point of time to control behavior and/or persist. Both allocation and selective stabilization resonate with the conceptual framework of Darwinian competition.

Observational and tagging experimental studies agree with computational theories [e.g., (*144*)] that an engram is sparsely encoded. That is, not all neurons within a given brain region become an engram cell supporting a particular memory. The size of an engram within a given brain region (that is, the number of engram cells) is stable and invariant to memory strength. For instance, the size of an LA engram (number of LA engram cells) is similar for an auditory fear conditioned memory and a cocaine-cue memory (*66*, *122*, *145*), and memory strength does not affect engram size (*89*, *122*) [for review, see (*146*)]. Rather, a stronger memory engages a greater number of synapses between engram cells (*88*).

Several lines of evidence suggest that one mechanism constraining engram size involves inhibitory neurons. Thus, inhibiting parvalbumin-containing interneurons in the basolateral amygdala complex increased the size of an engram in the LA supporting an auditory fear memory through a process involving disynaptic inhibition (*145*), in which an excitatory neuron inhibits another excitatory neuron via an intervening inhibitory neuron. Moreover, inhibiting somatostatin-containing interneurons increased the size of a DG contextual fear memory engram through a lateral-inhibition like process (*147*). The importance of inhibitory neurons in engrams has also been highlighted in human studies. For instance, evidence suggests that in the cortex, associative memories are represented in excitatory engrams and matched (equal and opposite) inhibitory engrams. Memories are expressed upon disinhibition of the excitatory engram (*148*–*150*). Further exploration of excitatory-inhibitory balance in engram formation, storage, and retrieval is necessary to understand how these opposing forces interact to support memory function.

Silent engrams in memory loss

Engrams may become damaged, such that a memory becomes forever unavailable. However, engrams may also be temporarily inaccessible, such that the engram still exists but cannot be retrieved by natural means. Silent engrams, engrams that cannot be retrieved by natural retrieval cues but can be retrieved with direct optogenetic stimulation, were first revealed in an experiment in which the protein synthesis inhibitor, anisomycin, was administered immediately after contextual fear conditioning in mice (*71*). Inhibiting protein synthesis before or immediately after an experience is known to induce amnesia (*151*, *152*) and block cellular consolidation (*153*–*155*). Cellular consolidation refers to the relatively fast process of memory stabilization thought to involve the expression of genes necessary to strengthen synapses. By contrast, systems consolidation (discussed below) refers to the slower, time-dependent reorganization of memories over distributed brain circuits (*156*–*159*).

In this study, mice administered anisomycin immediately after training showed little freezing when replaced in the training context 1 day later (*71*). Therefore, as expected, disrupting protein synthesis induced retrograde amnesia by blocking cellular consolidation. However, optogenetic reactivation of DG engram neurons tagged during contextual fear training was sufficient for memory recovery, even 8 days after training (Fig. 4). These results indicate that the engram was formed and persisted for several days but that this engram

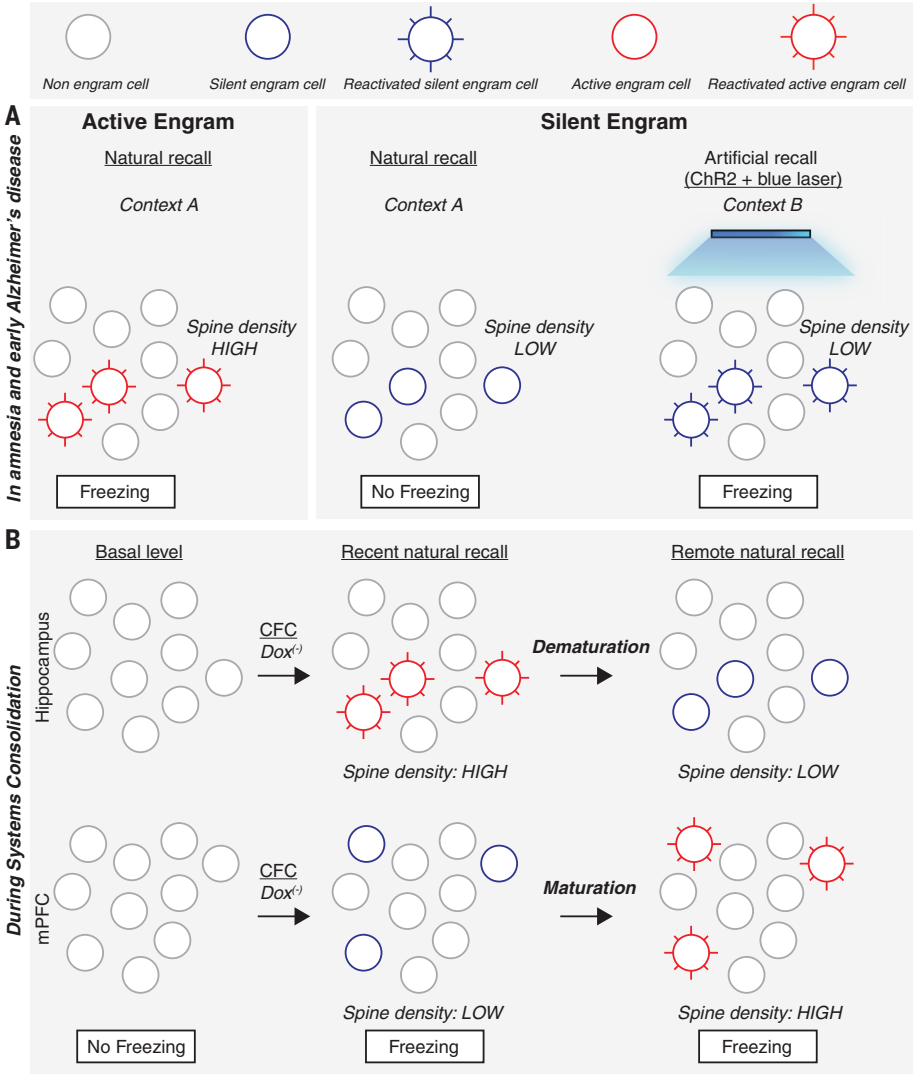


Fig. 4. Active and silent engram cells in amnesia and during memory systems consolidation. (**A**) Active engram cells have higher spine density and are activated in the conditioned context A to produce the conditioned response, freezing. Silent engram cells generated in amnesia and in a mouse designed to model early Alzheimer's disease show lower spine density and cannot be activated in the conditioned context A to produce a conditioned response but can be activated by blue light in an unconditioned context B if they were tagged with ChR2 during encoding. (**B**) During memory systems consolidation, active engram cells with high spine density are formed in the hippocampus during contextual fear conditioning and for several days, the conditioned context can evoke a conditioned response. However, by two weeks (remote recall), these hippocampal engram cells demature to become silent, with reduced spine density. In the mPFC, engram cells are formed during CFC but are silent with low spine density. During the following 2 weeks, these mPFC silent engram cells acquire higher spine density and become active engram cells.

could not be retrieved by natural means. Silent DG engram cells showed weaker physiological (increased synaptic strength) and structural (increased dendritic spine density) alterations than normal engram cells (in control mice), suggesting that a silent engram may be the result of disrupting the synaptic strengthening normally induced by training. That optogenetic activation of DG engram cells was able to induce memory retrieval suggests that direct optogenetic activation was able to circumvent this requirement for synaptic and structural plasticity within engram cells. Consistent with this, genetic restoration of spine density [targeted overexpression of p-21 activated kinase (PAK 1)] also allowed a silent engram to be reactivated and memory expressed by natural retrieval cues (*160*).

The idea that engrams may be silenced by disrupting synaptic efficacy and spine density and reawakened by enhancing synaptic plasticity is consistent with findings from a nonengram study examining auditory fear conditioning (*161*). Rats were trained in a variant of an auditory fear conditioning task in which the tone conditioned stimulus was replaced by optogenetic activation of LA axon terminals from neurons originating in the medial geniculate nucleus and auditory cortex. Immediately after conditioning, long-term depression (LTD)–like optogenetic stimulation was administered. LTD is thought to weaken synaptic efficacy and decrease spine density (*162*–*165*). Consistent with the interpretation that LTD-like stimulation silenced the engram, this opto-LTD stimulation impaired subsequent memory recall. However, LTP-like optogenetic stimulation allowed the memory to be retrieved (consistent with the interpretation that the engram was "unsilenced"). Again, subsequent LTD-like optogenetic stimulation silenced this memory, whereas LTP-like optogenetic stimulation allowed recovery of this memory.

These findings raise the question of whether engrams (and the memories they support) in other amnesic conditions are truly "lost" or are simply inaccessible such that they cannot be retrieved under natural conditions. Silent engrams were reactivated by artificially stimulating engram cells in amnestic mice used to study the early stages of Alzheimer's disease (AD) (*166*, *167*). These transgenic mice [APP/PS1 mice containing human transgenes with the familial AD mutation in both amyloid precursor protein (APP) and presenilin 1 (PSEN1)] showed contextual fear memory deficits (*166*). However, optogenetic reactivation of ChR2-labeled DG engram cells induced robust freezing comparable to control mice (*166*). Consistent with other examples of silent engram cells, DG engram cells in these mice used to study AD showed decreased spine density. However, LTP-like optogenetic stimulation at entorhinal cortex engram cell inputs onto DG engram cells restored not only spine density in DG engram cells but also the ability of natural retrieval cues to elicit memory retrieval (thereby unsilencing the engram) (*166*). These findings in mice are consistent with reports that memory retrieval in people with early-stage AD may be enhanced by particular retrieval cues (*168*, *169*). Therefore, under certain conditions, a previously inaccessible memory may be retrieved in human AD, consistent with the interpretation that some engrams in early-AD brains may be silent rather than lost.

Apart from clinical implications, the finding of silent engrams is relevant to discussions on the role of protein synthesis–dependent cellular consolidation in terms of memory storage versus retrieval. There has been persistent debate on this issue (*170*–*173*). The majority of neuroscientists examining cellular memory consolidation may favor the view that disrupting protein synthesis disrupts memory storage. However, in many amnesia experiments, memory storage is conflated with memory retrieval.

The finding that optogenetically stimulating a silent engram in an otherwise amnestic mouse, even 1 week after training, induces memory retrieval challenges the view that protein synthesis–dependent cellular consolidation is important for memory storage. Instead, these findings suggest that the role of cellular consolidation is to enhance subsequent retrievability of an engram, consistent with the idea of engram "retrieval handles" that are established after memory formation and may be remodeled after memory retrieval (*1*). Importantly, silent engrams are consistent with the pioneering cognitive psychologist Endel Tulving's (*174*) conceptual distinction between memory availability and accessibility, in which memory failure may reflect the absence of the information or difficulties accessing the information [see (*175*) for review].

Silent engrams in normal memory

Memory may change with time and circumstance. Might these changes in memory be mediated by endogenous engram silencing? This was explored in a social discrimination task in which mice interact more with a new, rather than a familiar, mouse. This social discrimination memory lasts roughly an hour after exposure to a familiar mouse (the training experience) and is absent 24 hours after training (*176*). The dorsal CA2 to ventral CA1 (vCA1) hippocampal circuit plays a pivotal role in social discrimination (*177*), with a vCA1 engram representing the familiar mouse (*178*). Consistent with the time course of social discrimination memory, the familiar mouse engram in vCA1 becomes silent an hour after training. However, artificially reactivating this engram 24 hours after training (when the social discrimination memory normally has dissipated) reinstates social discrimination memory, as if the trained-but-forgotten familiar mouse is being remembered. Besides artificial engram reactivation, the accessibility of vCA1 engram (and social discrimination memory) is prolonged by interventions such as group housing. These findings provide a hint that engram silencing may be one way in which the brain normally regulates mnemonic processes.

Additional evidence comes from memory extinction studies. After conditioning, repeated presentation of the conditioned stimuli alone (in the absence of the unconditioned stimulus) produces a gradual decrease of the conditioned response (*82*)—a phenomenon called extinction. Therefore, after extinction training, the ability of the conditioned stimulus to induce memory retrieval is diminished, an outcome that is similar phenomenologically to engram silencing. Might engram silencing account for extinction? Consistent with this general idea, some auditory fear extinction protocols induce synaptic depotentiation of LA neurons, that is, the reversal of synaptic potentiation induced by fear conditioning (*179*, *180*). Moreover, after fear conditioning, LTD-like electrical stimulation of external capsule inputs to the LA induces synaptic depotentiation and decreases fear behavior (*181*), resembling both extinction and engram silencing. Finally, shortly after extinction training, the chemogenetic artificial activation of cells tagged brainwide during context fear training (the putative fear engram) was reported to increase freezing levels (*182*), suggesting that the original fear engram was silenced during extinction. The similarities between engram silencing and extinction are consistent with theoretical views that during extinction, the conditioned stimulus–unconditioned stimulus contingency is "unlearned" (*183*, *184*).

However, other accounts stress that extinction does not reflect unlearning the original association (perhaps by silencing the original engram) but rather reflects learning a new "conditioned stimulus–no unconditioned stimulus" association (*185*, *186*) with a corresponding new extinction engram. That the original memory is not "erased" by extinction is suggested by findings that after extinction training, the conditioned response may return if the conditioned stimulus is presented (i) in a new nonextinction context (renewal), (ii) after a stressor (reinstatement), or (iii) after the passage of time (spontaneous recovery) (*187*–*192*). A recent study concluded that contextual fear extinction may be supported by a novel fear extinction engram in the DG that is distinct from and suppresses the contextual fear DG engram with a time course that corresponds to the emergence of spontaneous recovery (*53*). In this experiment, spontaneous recovery was observed remotely (29 days), but not recently (6 days), after extinction training. Moreover, the original fear engram was reactivated at the remote, but not recent, memory test after extinction training. The opposite pattern of results was observed for active cells tagged after extinction training (the presumed fear extinction engram). Interestingly, artificial reactivation of the fear extinction engram prevented spontaneous recovery of the original fear memory, even at remote times. These results suggest that the original fear engram and the extinction engram compete for control over behavior; the extinction engram first suppressed or silenced the original fear engram, but, with time, the fear extinction engram was itself silenced. Conversely, activation of a remote DG contextual fear engram (labeled 25 days after contextual fear conditioning) itself may also be important for subsequent fear memory extinction (*52*), perhaps similar to a process referred to as reconsolidation-updating (*193*, *194*). However, the extent to which DG neurons that were activated 25 days after contextual fear conditioning overlap with DG neurons active during training remains an open question (*40*, *51*).

Finally, a recent study examined fear extinction engrams in the amygdala and found that extinction engram cells were formed in a genetically distinct and "reward-responsive" subpopulation of basal amygdala neurons. These fear extinction engram cells suppressed the fear engram neurons that were also present in basal amygdala and, furthermore, induced appetitive behavior when optogenetically stimulated (*195*). These findings in mice are consistent with the results of a recent study in fruit flies (*26*) and highlight the similarities between fear extinction and reward processes across species. Moreover, these results are consistent with the general idea of competition between memory traces in the control of behavior.

Silent engrams and time

The representation of a memory in the brain may change with time. For instance, dorsal hippocampal lesions in rodents disrupt expression of contextual fear memories in the days, but not weeks after training (*196*–*198*). At more remote times, cortical areas, including anterior cingulate cortex or medial prefrontal cortex (mPFC), become preferentially engaged (*100*). The time-dependent reorganization of memory reflects systems consolidation, a process that typically refers to initially hippocampal-based episodic-like memories (*158*, *159*). Systems consolidation was recently examined at the level of the engram in the hippocampus and mPFC, where findings indicate time-dependent silencing of active engrams and conversions of silent engrams to active engrams (*51*, *199*). During contextual fear conditioning, active mPFC neurons were labeled to express ChR2. When placed in the conditioning context, mice showed robust freezing when tested either 2 days or 13 days after training. However, the engram ensemble components supporting memory retrieval differed with test time. Tagged mPFC neurons were reactivated 13 days, but not 2 days, after training, suggesting that the mPFC engram was silent shortly after training but active after longer delays. DG engram cells showed an opposite pattern; DG engram cells were reactivated shortly after training but silenced more remotely. Similar to other instances of silent engrams discussed above, the mPFC engram cells shortly after training and the DG engram cells at longer delays after training showed reduced spine density, and, furthermore, optogenetic activation of these silent engrams was sufficient to induce memory retrieval. Interestingly, posttraining tetanus toxin–induced inhibition of the input from DG engram cells to mPFC engram cells blocked the maturation of the silent mPFC engram cells to an accessible state, suggesting coordinated network function

between different engram ensemble components is important in systems consolidation.

Memories may also become less precise and more generalized with time (*200*–*202*). According to memory transformation theory, changes in the nature and quality of memories correspond to changes in neural representations, with hippocampal-dependent context-specific detailed memories transforming into gist-like schematic memories represented in cortical structures over time (*201*, *203*, *204*). The neural processes governing remote memory generalization at the engram level suggest that the availability of the DG engram is critical for memory specificity (*205*). In this experiment, shortly after contextual fear conditioning (1 day), mice froze in the training context only, whereas at more remote time points (16 days after training), mice also froze in a nonshocked context. This finding is consistent with previous reports of contextual fear memory generalizing over time (*51*, *201*). At the recent, but not remote, time, DG engram cells showed greater connectivity to parvalbumin-expressing CA3 basket cells (thereby inhibiting CA3 pyramidal neurons through feedforward inhibition) than nonengram DG cells, suggesting that greater feedforward inhibition in DG-CA3 circuits helps maintain memory precision. Interestingly, optogenetic activation of DG engram cells 10 days after training did not induce memory retrieval (suggesting that this engram had become unavailable), except if feedforward inhibition of CA3 pyramidal neurons was genetically enhanced. Moreover, mice with genetically enhanced feedforward inhibition also showed precise memory, even when tested at more remote times. Together, these data suggest that enhanced feedforward inhibition onto CA3 neurons maintains DG engram cell availability and delays the loss of context specificity associated with remote memories.

These findings suggest that engram silencing may represent a continuum of a natural state of an engram. That is, an engram may be (i) unavailable (neither natural conditioned stimuli nor artificial reactivation induces memory expression), (ii) silenced (only artificial reactivation is sufficient to induce memory expression), (iii) dormant or latent, as initially named by Semon (natural conditioned stimuli may induce memory retrieval), or (iv) active (currently being retrieved). Different processes may mediate these distinct engram states. For example, similar to silencing a DG engram, posttraining anisomycin administration silenced an LA engram supporting an auditory fear memory (*79*). However, if in addition to anisomycin, a peptide to induce autophagy (a mechanism of protein degradation) was administered after training, then optogenetic reactivation of inputs to the LA was no longer sufficient to induce memory retrieval (*206*), suggesting that autophagy made the engram unavailable rather than simply inaccessible.

From engrams to knowledge

Thus far, we have discussed engrams supporting a single memory. Of course, animals (including humans) learn and remember many things. Some of these experiences may be best remembered as distinct episodes, rich with episodic details (*207*–*209*). However, in other circumstances, it may be advantageous to link related experiences, thereby creating a general concept or principle (*210*–*214*). This raises the question of how engrams representing different experiences interact. The mechanisms governing neuronal allocation to an engram supporting a single experience also serve to either coallocate neurons to overlapping engrams (thereby linking experiences) or disallocate neurons to nonoverlapping engrams (thereby disambiguating experiences) (*121*, *215*–*217*) (Fig. 5). In this way, relative neuronal excitability is critical not only for initial engram formation but also in organizing different memory representations across the brain.

Neurons that are relatively more excitable than their neighbors at the time of an experience are more likely to be allocated to the engram supporting the memory of that experience (*121*). Increased excitability in engram cells is also maintained for several hours after an experience (*215*, *218*, *219*). Therefore, if a related experience occurs in this time window, these same (or overlapping) engram cells are more excitable than their neighbors and thus coallocated to the engram supporting the memory of the second experience. Because the memories of the two experiences are coallocated to overlapping engram cells, these two memories become linked (or integrated); thinking of one experience automatically makes one think of the second. For example, LA neurons allocated to one fear memory were coallocated to a second fear memory if the second event occurred minutes to hours (30 min to 6 hours), but not 24 hours, after the first (*215*). This linking occurred even if the conditioned stimuli used in the two training sessions were of different modalities (e.g., a light and a tone or a context and a tone). Similarly, coallocation of CA1 engram cells supporting memories of two distinct contexts was observed if exposure to the contexts was separated by a short time interval (*216*). Behaviorally extinguishing one memory produced extinction for the second memory, even though the second memory was not behaviorally extinguished, indicating that the two memories were functionally linked (*215*). Coallocated memories may maintain their distinct identity by engaging specific synapses within shared engram cells (*79*). Moreover, in addition to integrating two similar memories (two fear memories or two contextual memories), two aversive, but otherwise dissimilar memories (a conditioned fear and a conditioned taste aversion memory), were integrated by repeated coretrieval of these memories (*220*). Overall, these data from rodent experiments agree with results from human memory experiments showing that the representations of memories for events experienced close in time or with related content overlap may be integrated or linked, thus enabling generalization and flexible use of this shared information [e.g., (*212*, *221*–*224*)].

Memory retrieval also transiently reactivates engram cells (*89*, *215*, *219*). This increase in excitability both enhances the precision and efficiency of memory retrieval (*219*) and opens a new "coallocation window" (*215*), perhaps explaining how new information is integrated into preexisting knowledge.

Conclusions and perspectives

Overall, these studies provide persuasive evidence for the existence of engrams in rodent brains. We agree with Endel Tulving who stated "As a scientist I am compelled to the conclusion—not postulation, not assumption, but conclusion—that there must exist certain physical-chemical changes in the nervous tissue that correspond to the storage of information, or to the engram, changes that constitute the necessary conditions of remembering. (The alternative stance, that it may be possible for any behavior or any thought to occur independently of physical changes in the nervous system, as all your good readers know, is sheer mysticism)" (*225*). The findings from many labs using different methods to examine many types of memory converge to support the idea that complex information may not be represented in single cells [e.g., a "grandmother cell" (*226*, *227*)]; instead, these findings suggest that the basic unit of computation in the brain is an engram (*228*, *229*).

To understand a complex, multilayered system such as the brain, it is crucial to causally link a process or phenomenon occurring at a lower level of complexity to those at higher levels. Traditionally, such studies have been carried out using interventions such as tissue lesion or pharmacological disruption. Many of the studies discussed in this review took advantage of state-of-the-art intervention techniques and their combinations, including temporally inducible targeted transgenics and optogenetics, that may generally permit the identification of more precise cause-consequence relationships. Nevertheless, even advanced interventions inevitably artificially manipulate the brain and therefore provide information as to what an engram can do, but not necessarily what it does do (physiologically). This point has been articulated in several other reviews on memory research [e.g., (*230*)].

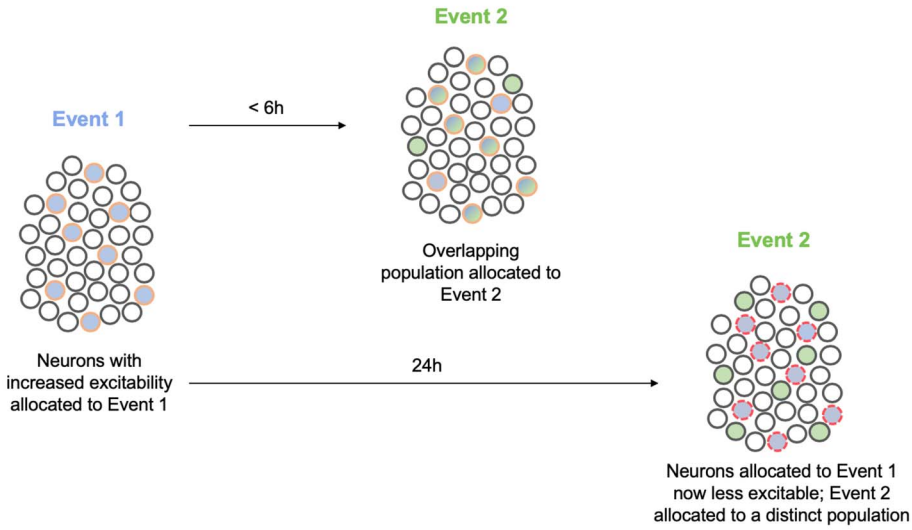


Fig. 5. Neuronal allocation and memory linking. Neurons with increased excitability at the time of event 1 (blue) are allocated to the engram supporting this memory (blue filled circles outlined in orange). These allocated engram neurons remain more excitable than their neighbors for several hours after event 1. If a similar event 2 (green) occurs during this time, neurons allocated to the engram supporting event 1 are more excitable and, therefore, also allocated to the engram supporting event 2 (blue and green filled circles outlined in orange). In this way, neurons are coallocated to events 1 and 2. By virtue of coallocation, these two memories become linked. After some time, neurons allocated to the engram supporting event 1 become less excitable than their neighbors ("refractory"), and if event 2 occurs in this time window, a new population of more excitable neurons wins the competition for allocation to the engram supporting event 2. This disallocation allows the two memories to be remembered separately. Circles with red dashed outlines represent less excitable neurons.

However, the results of these intervention studies provide direction as to which processes we should focus our efforts to understand how the brain actually forms and retrieves memory. Furthermore, the high specificity of the state-of-the-art intervention methods, spanning from the molecular level up to the behavioral level, have already revealed mechanisms that would have been difficult to study using other techniques. For instance, these artificial intervention studies allowed the field to identify the silent state of an engram and the mechanism underlying memory allocation.

More than 100 years ago, Semon put forth a law of engraphy. Combining these theoretical ideas with the new tools that allow researchers to image and manipulate engrams at the level of cell ensembles facilitated many important insights into memory function. For instance, evidence indicates that both increased intrinsic excitability and synaptic plasticity work hand in hand to form engrams and that these processes may also be important in memory linking, memory retrieval, and memory consolidation. Interestingly, disrupting synaptic plasticity in engram cells either by disease processes (as in mice used to study AD) or amnestic drugs (such as protein synthesis inhibitors) or during some natural behaviors (housing condition in social discrimination memory, memory systems consolidation, and perhaps fear extinction training) silences engrams such that they can no longer be accessed by normal sensory cues. However, these studies show that silent engrams still exist in the brain and that the information they represent may not be forever lost. The pioneering psychologist and behaviorist Edward Tolman (*231*) advanced the concept of latent learning and latent memory: learning that occurs without reinforcement, the memory of which is not revealed or expressed until the need or motivation for the acquired knowledge arises (*232*, *233*). It would be interesting to determine whether at least some latent memories are based on silent engrams and, if so, use the conversion of silent engram to active engram as a means of identifying and characterizing the brain circuits mediating the relevant motivation.

A continuum of engram accessibility states may exist. Engrams may be entirely unavailable and not retrievable, even through artificial means (the memory would be forgotten). Or, engrams may be silenced such that memories may be retrieved by artificially reactivating engram cells. The processes that silence or erase an engram, as well as strategies for unsilencing engrams, are a subject for further investigation. That it was possible to artificially reactivate silent engrams in mice designed to study the memory deficits of AD hint at the extraordinary translational potential of this line of research.

Some additional general themes emerge from the results of engram studies. The first theme is that findings from engram studies are reminiscent of reconsolidation studies. Upon retrieval, a memory may enter a labile and modifiable state that lasts for several hours. The process of restabilizing this memory is referred to as reconsolidation. Although reconsolidation has a longer history (*234*), the modern reawakening of this phenomenon stems from a finding by Nader, LeDoux, and Schafe (*235*). At the time that this groundbreaking study was conducted, the general thinking was that memories become stabilized in a process of cellular consolidation that occurs once, shortly after a learning experience. However, Nader, LeDoux, and Schafe challenged this view by showing that memory retrieval opens a several-hour "reconsolidation window" during which different interventions may weaken or strengthen the original memory. For instance, disrupting protein synthesis during the reconsolidation window of a conditioned fear memory produced apparent amnesia for this memory. This result was replicated and generalized to several types of memory (*156*, *236*, *237*). There are many similarities between this reconsolidation blockade and engram silencing. For instance, reconsolidation blockade is only observed when a memory is being actively retrieved, because administering anisomycin (or another similar intervention) in the absence of memory reactivation does not impair its subsequent retrieval. Viewed from an "engram conceptual framework," retrieval of a specific memory would activate the underlying engram, and disrupting protein synthesis shortly after this activation might silence this engram. The function of reconsolidation may be to update a memory (*1*, *211*, *238*–*240*). That the reconsolidation window is not unlike the coallocation window suggests that these two processes might be similar ways of explaining the same (or similar) phenomenon at different levels of analysis.

A second emerging theme is that of competition. Allocation to an engram involves competition between eligible neurons within a given brain region at the time of memory encoding. Competition represents a fundamental property of many biological systems and has been previously shown to be important in other mnemonic phenomena. For instance, memory traces may compete for control of behavior at the time of retrieval (*241*). In addition, human studies reveal that memories may compete if they are linked to a common retrieval cue. Retrieval of a target memory may lead to retrieval-induced forgetting of currently irrelevant competing memories (*242*).

Although recent engram studies have offered important insights into memory, several key questions remain. First, although the majority of observational studies reveal that the overlap between populations of neurons active during training and testing exceed chance levels, the overall correspondence between these two populations is relatively low (roughly 10 to 40%, depending on the study). That this overlap does not approach 100% suggests a number of possibilities. First, the methods to label active neurons using IEG promoters may be imprecise (either "overtagging" or "undertagging" the "real" engram at training and/or testing). Alternatively, engrams may be dynamic, even over relatively short (days) periods of time, with cells "dropping into" or "dropping out of" the engram as it is refined or consolidated (*243*, *244*). It will be interesting to determine how the mechanisms of engram silencing contribute to and/or interact with this refinement process and the implications this may have on memory quality, precision, or strength. Moreover, it will be important to determine how engrams change over more prolonged periods of time. For example, do all engrams (engrams representing different types of memories such as episodic, semantic, or even procedural or motor memories, with different valence) change over time, gradually engaging more cortical regions? Is there a role for top-down (mPFC to hippocampal) processing in the dematuration of hippocampal engrams and a possible role of silent hippocampal engrams in remote memory recall?

Second, how can we leverage our knowledge of engrams in rodents to better understand human memory? There is good evidence for general engram-like memory representations in humans [e.g., (*245*)], but, to date, there are no compelling findings at the cellular ensemble level. To extend the findings from rodent engram studies to humans, it may be necessary to develop non- to low-invasive methods to image and manipulate engrams at the single-cell or specific ensemble level in humans. Progress in this general area of human "artificial memory manipulation" has been made by harnessing the power of reconsolidation (*194*, *235*, *237*, *246*, *247*) in which engram cells are thought to be specifically reactivated by memory retrieval. Pharmacological blockade of reconsolidation and noninvasive techniques that "update" memory during reconsolidation have shown some success in manipulating human memories (*248*, *249*).

Finally, it is important that the links between neuroscience and artificial intelligence (AI) are leveraged to inform both fields. Understanding how the brain encodes, stores, and uses information, especially at the level of the engram, can help inspire the development of more intelligent machines. For instance, engrams and how engrams serve to link memories and organize information in the brain may motivate the development of new algorithms and AI architectures to better allow these agents to form generalizations and schema. In addition, machine learning and deep neural networks may inspire or generate testable theories at the level of the engram for neuroscientists to investigate. In this way, uniting the foundational theories of AI pioneer Alan Turing with those of Endel Tulving could benefit both AI and memory research.

REFERENCES AND NOTES

1. Y. Dudai, The neurobiology of consolidations, or, how stable is the engram? *Annu. Rev. Psychol.* **55**, 51–86 (2004). doi: 10.1146/annurev.psych.55.090902.142050; pmid: 14744210
2. D. L. Schacter, Constructive memory: Past and future. *Dialogues Clin. Neurosci.* **14**, 7–18 (2012). pmid: 22577300
3. D. L. Schacter, *Stranger Behind the Engram: Theories of Memory and the Psychology of Science* (Erlbaum Associates, 1982).
4. D. L. Schacter, J. E. Eich, E. Tulving, Richard Semon's theory of memory. *J. Verbal Learn. Verbal Behav.* **17**, 721–743 (1978). doi: 10.1016/S0022-5371(78)90443-7
5. R. Semon, *The Mneme* (G. Allen & Unwin, 1921).
6. R. Semon, *Die Mneme als erhaltendes Prinzip im Wechsel des organischen Geschehens*, W. Engelmann, Ed. (Leipzig, 1904).
7. R. W. Semon, *Mnemic Psychology* (G. Allen & Unwin, 1923).
8. S. A. Josselyn, S. Köhler, P. W. Frankland, Finding the engram. *Nat. Rev. Neurosci.* **16**, 521–534 (2015). doi: 10.1038/nrn4000; pmid: 26289572
9. S. A. Josselyn, S. Köhler, P. W. Frankland, Heroes of the engram. *J. Neurosci.* **37**, 4647–4657 (2017). doi: 10.1523/JNEUROSCI.0056-17.2017; pmid: 28469009
10. S. Tonegawa, X. Liu, S. Ramirez, R. Redondo, Memory engram cells have come of age. *Neuron* **87**, 918–931 (2015). doi: 10.1016/j.neuron.2015.08.002; pmid: 26335640
11. D. L. Schacter, *Forgotten Ideas, Neglected Pioneers: Richard Semon and the Story of Memory* (Psychology Press, 2001).
12. K. S. Lashley, in *Society of Experimental Biology Symposium, No. 4: Psychological Mechanisms in Animal Behavior*, J. F. Danielli, R. Brown, Eds. (Academic Press, 1950), pp. 454–482.
13. K. S. Lashley, *Brain Mechanisms and Intelligence: A Quantitative Study of Injuries to the Brain* (Dover Books on Psychology, vol. T1038, Dover Publications, 1963).
14. K. S. Lashley, Mass action in cerebral function. *Science* **73**, 245–254 (1931). doi: 10.1126/science.73.1888.245; pmid: 17755301
15. D. O. Hebb, *The Organization of Behavior: A Neuropsychological Theory* (Wiley, 1949).
16. C. J. Shatz, The developing brain. *Sci. Am.* **267**, 60–67 (1992). doi: 10.1038/scientificamerican0992-60; pmid: 1502524
17. D. Marr, Simple memory: A theory for archicortex. *Philos. Trans. R. Soc. Lond. Ser. B* **262**, 23–81 (1971). doi: 10.1098/rstb.1971.0078; pmid: 4399412
18. M. R. Hunsaker, R. P. Kesner, The operation of pattern separation and pattern completion processes associated with different attributes or domains of memory. *Neurosci. Biobehav. Rev.* **37**, 36–58 (2013). doi: 10.1016/j.neubiorev.2012.09.014; pmid: 23043857
19. J. J. Knierim, J. P. Neunuebel, Tracking the flow of hippocampal computation: Pattern separation, pattern completion, and attractor dynamics. *Neurobiol. Learn. Mem.* **129**, 38–49 (2016). doi: 10.1016/j.nlm.2015.10.008; pmid: 26514299
20. M. Moscovitch, in *Science of Memory: Concepts*, H. L. I. Roediger, Y. Dudai, S. M. Fitzpatrick, Eds. (Oxford Univ. Press, 2007), pp. 17–29.
21. S. J. Martin, P. D. Grimwood, R. G. Morris, Synaptic plasticity and memory: An evaluation of the hypothesis. *Annu. Rev. Neurosci.* **23**, 649–711 (2000). doi: 10.1146/annurev.neuro.23.1.649; pmid: 10845078
22. S. J. Martin, R. G. Morris, New life in an old idea: The synaptic plasticity and memory hypothesis revisited. *Hippocampus* **12**, 609–636 (2002). doi: 10.1002/hipo.10107; pmid: 12440577
23. D. Yu, Y. Tan, M. Chakraborty, S. Tomchik, R. L. Davis, Elongator complex is required for long-term olfactory memory formation in *Drosophila*. *Learn. Mem.* **25**, 183–196 (2018). doi: 10.1101/lm.046557.117; pmid: 29545390
24. D. Owald *et al.*, Activity of defined mushroom body output neurons underlies learned olfactory behavior in *Drosophila*. *Neuron* **86**, 417–427 (2015). doi: 10.1016/j.neuron.2015.03.025; pmid: 25864636
25. E. Perisse *et al.*, Aversive learning and appetitive motivation toggle feed-forward inhibition in the *Drosophila* mushroom body. *Neuron* **90**, 1086–1099 (2016). doi: 10.1016/j.neuron.2016.04.034; pmid: 27210550
26. J. Felsenberg *et al.*, Integration of parallel opposing memories underlies memory extinction. *Cell* **175**, 709–722.e15 (2018). doi: 10.1016/j.cell.2018.08.021; pmid: 30245010
27. T. Miyashita, E. Kikuchi, J. Horiuchi, M. Saitoe, Long-term memory engram cells are established by c-Fos/CREB transcriptional cycling. *Cell Rep.* **25**, 2716–2728.e3 (2018). doi: 10.1016/j.celrep.2018.11.022; pmid: 30517860
28. J. Z. Young, Computation in the learning system of cephalopods. *Biol. Bull.* **180**, 200–208 (1991). doi: 10.2307/1542389; pmid: 29304688
29. J. Z. Young, in *Cephalopod Neurobiology: Neuroscience Studies in Squid, Octopus and Cuttlefish*, N. J. Abbott, R. Williamson, L. Maddock, Eds. (Oxford Univ. Press, 1995), pp. 431–443.
30. A. Bédécarrats, S. Chen, K. Pearce, D. Cai, D. L. Glanzman, RNA from trained *Aplysia* can induce an epigenetic engram for long-term sensitization in untrained *Aplysia*. *eNeuro* **5**, ENEURO.0038-0018.2018 (2018). doi: 10.1523/ENEURO.0038-18.2018; pmid: 29789810
31. T. J. Carew, R. D. Hawkins, T. W. Abrams, E. R. Kandel, A test of Hebb's postulate at identified synapses which mediate classical conditioning in *Aplysia*. *J. Neurosci.* **4**, 1217–1224 (1984). doi: 10.1523/JNEUROSCI.04-05-01217.1984; pmid: 6726327
32. R. Menzel, Searching for the memory trace in a mini-brain, the honeybee. *Learn. Mem.* **8**, 53–62 (2001). doi: 10.1101/lm.38801; pmid: 11274250
33. D. L. Alkon, Calcium-mediated reduction of ionic currents: A biophysical memory trace. *Science* **226**, 1037–1045 (1984). doi: 10.1126/science.6093258; pmid: 6093258
34. D. A. Mccormick *et al.*, The engram found? Role of the cerebellum in classical conditioning of nictitating membrane and eyelid responses. *Bull. Psychon. Soc.* **18**, 103–105 (1981). doi: 10.3758/BF03333573
35. J. M. Fuster, J. P. Jervey, Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science* **212**, 952–955 (1981). doi: 10.1126/science.7233192; pmid: 7233192
36. Y. Miyashita, Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature* **335**, 817–820 (1988). doi: 10.1038/335817a0; pmid: 3185711
37. M. E. Greenberg, E. B. Ziff, Stimulation of 3T3 cells induces transcription of the *c-fos* proto-oncogene. *Nature* **311**, 433–438 (1984). doi: 10.1038/311433a0; pmid: 6090941
38. T. Curran, J. I. Morgan, Superinduction of *c-fos* by nerve growth factor in the presence of peripherally active benzodiazepines. *Science* **229**, 1265–1268 (1985). doi: 10.1126/science.4035354; pmid: 4035354
39. J. F. Guzowski, B. L. McNaughton, C. A. Barnes, P. F. Worley, Imaging neural activity with temporal and cellular resolution using FISH. *Curr. Opin. Neurobiol.* **11**, 579–584 (2001). doi: 10.1016/S0959-4388(00)00252-X; pmid: 11595491
40. C. A. Denny *et al.*, Hippocampal memory traces are differentially modulated by experience, time, and adult neurogenesis. *Neuron* **83**, 189–201 (2014). doi: 10.1016/j.neuron.2014.05.018; pmid: 24991962
41. L. G. Reijmers, B. L. Perkins, N. Matsuo, M. Mayford, Localization of a stable neural correlate of associative memory. *Science* **317**, 1230–1233 (2007). doi: 10.1126/science.1143839; pmid: 17761885
42. L. A. DeNardo *et al.*, Temporal evolution of cortical ensembles promoting remote memory retrieval. *Nat. Neurosci.* **22**, 460–469 (2019). doi: 10.1038/s41593-018-0318-7; pmid: 30692687
43. A. T. Sørensen *et al.*, A robust activity marking system for exploring active neuronal ensembles. *eLife* **5**, e13918 (2016). doi: 10.7554/eLife.13918; pmid: 27661450
44. M. S. Fanselow, L. S. Lester, in *Evolution and Learning*, R. C. Bolles, M. D. Beecher, Eds. (Lawrence Erlbaum Associates, Inc., 1988), pp. 185–212.
45. K. K. Tayler, K. Z. Tanaka, L. G. Reijmers, B. J. Wiltgen, Reactivation of neural ensembles during the retrieval of recent and remote memory. *Curr. Biol.* **23**, 99–106 (2013). doi: 10.1016/j.cub.2012.11.019; pmid: 23246402

46. S. Ramirez *et al*., Creating a false memory in the hippocampus. *Science* **341**, 387–391 (2013). doi: 10.1126/science.1239073; pmid: 23888038
47. W. Deng, M. Mayford, F. H. Gage, Selection of distinct populations of dentate granule cells in response to inputs as a mechanism for pattern separation in mice. *eLife* **2**, e00312 (2013). doi: 10.7554/eLife.00312; pmid: 23538967
48. K. Z. Tanaka *et al*., Cortical representations are reinstated by the hippocampus during memory retrieval. *Neuron* **84**, 347–354 (2014). doi: 10.1016/j.neuron.2014.09.037 pmid: 25308331
49. M. Zelikowsky, S. Hersman, M. K. Chawla, C. A. Barnes, M. S. Fanselow, Neuronal ensembles in amygdala, hippocampus, and prefrontal cortex track differential components of contextual fear. *J. Neurosci.* **34**, 8462–8466 (2014). doi: 10.1523/JNEUROSCI.3624-13.2014; pmid: 24948801
50. Y. Nakazawa, A. Pevzner, K. Z. Tanaka, B. J. Wiltgen, Memory retrieval along the proximodistal axis of CA1. *Hippocampus* **26**, 1140–1148 (2016). doi: 10.1002/hipo.22596; pmid: 27068122
51. T. Kitamura *et al*., Engrams and circuits crucial for systems consolidation of a memory. *Science* **356**, 73–78 (2017). doi: 10.1126/science.aam6808; pmid: 28386011
52. O. Khalaf *et al*., Reactivation of recall-induced neurons contributes to remote fear memory attenuation. *Science* **360**, 1239–1242 (2018). doi: 10.1126/science.aas9875 pmid: 29903974
53. A. F. Lacagnina *et al*., Distinct hippocampal engrams control extinction and relapse of fear memory. *Nat. Neurosci.* **22**, 753–761 (2019). doi: 10.1038/s41593-019-0361-z; pmid: 30936555
54. H. Nomura, C. Teshirogi, D. Nakayama, M. Minami, Y. Ikegaya, Prior observation of fear learning enhances subsequent self-experienced fear learning with an overlapping neuronal ensemble in the dorsal hippocampus. *Mol. Brain* **12**, 21 (2019). doi: 10.1186/s13041-019-0443-6; pmid: 30871580
55. S. Trouche, J. M. Sasaki, T. Tu, L. G. Reijmers, Fear extinction causes target-specific remodeling of perisomatic inhibitory synapses. *Neuron* **80**, 1054–1065 (2013). doi: 10.1016/j.neuron.2013.07.047; pmid: 24183705
56. A. Nonaka *et al*., Synaptic plasticity associated with a memory engram in the basolateral amygdala. *J. Neurosci.* **34**, 9305–9309 (2014). doi: 10.1523/JNEUROSCI.4233-13.2014; pmid: 25009263
57. H. Xie *et al*., In vivo imaging of immediate early gene expression reveals layer-specific memory traces in the mammalian brain. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2788–2793 (2014). doi: 10.1073/pnas.1316808111 pmid: 24550309
58. J. H. Han *et al*., Selective erasure of a fear memory. *Science* **323**, 1492–1496 (2009). doi: 10.1126/science.1164139; pmid: 19286560
59. Y. Dong *et al*., CREB modulates excitability of nucleus accumbens neurons. *Nat. Neurosci.* **9**, 475–477 (2006). doi: 10.1038/nn1661; pmid: 16520736
60. H. Marie, W. Morishita, X. Yu, N. Calakos, R. C. Malenka, Generation of silent synapses by acute in vivo expression of CaMKIV and CREB. *Neuron* **45**, 741–752 (2005). doi: 10.1016/j.neuron.2005.01.039; pmid: 15748849
61. M. H. Han *et al*., Role of cAMP response element-binding protein in the rat locus ceruleus: Regulation of neuronal activity and opiate withdrawal behaviors. *J. Neurosci.* **26**, 4624–4629 (2006). doi: 10.1523/JNEUROSCI.4701-05.2006; pmid: 16641242
62. M. Lopez de Armentia *et al*., cAMP response element-binding protein-mediated gene expression increases the intrinsic excitability of CA1 pyramidal neurons. *J. Neurosci.* **27**, 13909–13918 (2016). doi: 10.1523/JNEUROSCI.3850-07.2007; pmid: 18077703
63. Y. Zhou *et al*., CREB regulates excitability and the allocation of memory to subsets of neurons in the amygdala. *Nat. Neurosci.* **12**, 1438–1443 (2009). doi: 10.1038/nn.2405; pmid: 19783993
64. E. Benito, A. Barco, CREB's control of intrinsic and synaptic plasticity: Implications for CREB-dependent memory models. *Trends Neurosci.* **33**, 230–240 (2010). doi: 10.1016/j.tins.2010.02.001; pmid: 20223527
65. D. Sargin *et al*., CREB regulates spine density of lateral amygdala neurons: Implications for memory allocation. *Front. Behav. Neurosci.* **7**, 209 (2013). doi: 10.3389/fnbeh.2013.00209; pmid: 24391565
66. H. L. Hsiang *et al*., Manipulating a "cocaine engram" in mice. *J. Neurosci.* **34**, 14115–14127 (2014). doi: 10.1523/JNEUROSCI.3327-14.2014; pmid: 25319707
67. E. Koya *et al*., Targeted disruption of cocaine-activated nucleus accumbens neurons prevents context-specific sensitization. *Nat. Neurosci.* **12**, 1069–1073 (2009). doi: 10.1038/nn.2364; pmid: 19620976
68. P. M. Milner, Cell assemblies: Whose idea? *Psycholoquy* **10**, 1 (1999).
69. X. Liu *et al*., Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature* **484**, 381–385 (2012). doi: 10.1038/nature11028; pmid: 22441246
70. E. S. Boyden, F. Zhang, E. Bamberg, G. Nagel, K. Deisseroth, Millisecond-timescale, genetically targeted optical control of neural activity. *Nat. Neurosci.* **8**, 1263–1268 (2005). doi: 10.1038/nn1525; pmid: 16116447
71. T. J. Ryan, D. S. Roy, M. Pignatelli, A. Arons, S. Tonegawa, Engram cells retain memory under retrograde amnesia. *Science* **348**, 1007–1013 (2015). doi: 10.1126/science.aaa5542; pmid: 26023136
72. B. N. Armbruster, X. Li, M. H. Pausch, S. Herlitze, B. L. Roth, Evolving the lock to fit the key to create a family of G protein-coupled receptors potently activated by an inert ligand. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5163–5168 (2007). doi: 10.1073/pnas.0700293104; pmid: 17360345
73. C. D. Nichols, B. L. Roth, Engineered G-protein coupled receptors are powerful tools to investigate biological processes and behaviors. *Front. Mol. Neurosci.* **2**, 16 (2009). doi: 10.3389/neuro.02.016.2009; pmid: 19893765
74. K. K. Cowansage *et al*., Direct reactivation of a coherent neocortical memory of context. *Neuron* **84**, 432–441 (2014). doi: 10.1016/j.neuron.2014.09.022; pmid: 25308330
75. R. L. Redondo *et al*., Bidirectional switch of the valence associated with a hippocampal contextual memory engram. *Nature* **513**, 426–430 (2014). doi: 10.1038/nature13725; pmid: 25162525
76. A. P. Yiu *et al*., Neurons are recruited to a memory trace based on relative neuronal excitability immediately before training. *Neuron* **83**, 722–735 (2014). doi: 10.1016/j.neuron.2014.07.017; pmid: 25102562
77. T. Rogerson *et al*., Synaptic tagging during memory allocation. *Nat. Rev. Neurosci.* **15**, 157–169 (2014). doi: 10.1038/nrn3667; pmid: 24496410
78. J. Kim, J. T. Kwon, H. S. Kim, S. A. Josselyn, J. H. Han, Memory recall and modifications by activating neurons with elevated CREB. *Nat. Neurosci.* **17**, 65–72 (2014). doi: 10.1038/nn.3592; pmid: 24212670
79. K. Abdou *et al*., Synapse-specific representation of the identity of overlapping memory engrams. *Science* **360**, 1227–1231 (2018). doi: 10.1126/science.aat3810 pmid: 29903972
80. A. Guskjolen *et al*., Recovery of "lost" infant memories in mice. *Curr. Biol.* **28**, 2283–2290.e3 (2018). doi: 10.1016/j.cub.2018.05.059; pmid: 29983316
81. K. Ghandour *et al*., Orchestrated ensemble activities constitute a hippocampal memory engram. *Nat. Commun.* **10**, 2637 (2019). doi: 10.1038/s41467-019-10683-2; pmid: 31201332
82. I. Pavlov, *Conditioned Reflexes* (Oxford Univ. Press, 1927).
83. D. L. Schacter, E. F. Loftus, Memory and law: What can cognitive neuroscience contribute? *Nat. Neurosci.* **16**, 119–123 (2013). doi: 10.1038/nn.3294; pmid: 23354384
84. A. R. Garner *et al*., Generation of a synthetic memory trace. *Science* **335**, 1513–1516 (2012). doi: 10.1126/science.1214985; pmid: 22442487
85. N. Ohkawa *et al*., Artificial association of pre-stored information to generate a qualitatively new memory. *Cell Rep.* **11**, 261–269 (2015). doi: 10.1016/j.celrep.2015.03.017; pmid: 25843716
86. G. Vetere *et al*., Memory formation in the absence of experience. *Nat. Neurosci.* **22**, 933–940 (2019). doi: 10.1038/s41593-019-0389-0; pmid: 31036944
87. J. A. Kauer, R. C. Malenka, Synaptic plasticity and addiction. *Nat. Rev. Neurosci.* **8**, 844–858 (2007). doi: 10.1038/nrn2234; pmid: 17948030
88. J. H. Choi *et al*., Interregional synaptic maps among engram cells underlie memory formation. *Science* **360**, 430–435 (2018). doi: 10.1126/science.aas9204; pmid: 29700265
89. L. A. Gouty-Colomer *et al*., *Arc* expression identifies the lateral amygdala fear memory trace. *Mol. Psychiatry* **21**, 364–375 (2016). doi: 10.1038/mp.2015.18; pmid: 25802982
90. A. Hayashi-Takagi *et al*., Labelling and optical erasure of synaptic memory traces in the motor cortex. *Nature* **525**, 333–338 (2015). doi: 10.1038/nature15257; pmid: 26352471
91. S. Park *et al*., Neuronal allocation to a hippocampal engram. *Neuropsychopharmacology* **41**, 2987–2993 (2016). doi: 10.1038/npp.2016.73; pmid: 27187069
92. M. J. Sekeres *et al*., Increasing CRTC1 function in the dentate gyrus during memory formation or reactivation increases memory strength without compromising memory quality. *J. Neurosci.* **32**, 17857–17868 (2012). doi: 10.1523/JNEUROSCI.1419-12.2012; pmid: 23223304
93. M. J. Sekeres, R. L. Neve, P. W. Frankland, S. A. Josselyn, Dorsal hippocampal CREB is both necessary and sufficient for spatial memory. *Learn. Mem.* **17**, 280–283 (2010). doi: 10.1101/lm.1785510; pmid: 20495061
94. T. Sacco, B. Sacchetti, Role of secondary sensory cortices in emotional memory storage and retrieval in rats. *Science* **329**, 649–656 (2010). doi: 10.1126/science.1183165; pmid: 20689011
95. W. B. Kim, J. H. Cho, Synaptic targeting of double-projecting ventral CA1 hippocampal neurons to the medial prefrontal cortex and basal amygdala. *J. Neurosci.* **37**, 4868–4882 (2017). doi: 10.1523/JNEUROSCI.3579-16.2017; pmid: 28385873
96. Y. Yang *et al*., Selective synaptic remodeling of amygdalocortical connections associated with fear memory. *Nat. Neurosci.* **19**, 1348–1355 (2016). doi: 10.1038/nn.4370; pmid: 27595384
97. A. L. Wheeler *et al*., Identification of a functional connectome for long-term fear memory in mice. *PLOS Comput. Biol.* **9**, e1002853 (2013). doi: 10.1371/journal.pcbi.1002853; pmid: 23300432
98. G. Vetere *et al*., Chemogenetic interrogation of a brain-wide fear memory network in mice. *Neuron* **94**, 363–374.e4 (2017). doi: 10.1016/j.neuron.2017.03.037; pmid: 28426969
99. D. S. Roy *et al*., Brain-wide mapping of contextual fear memory engram ensembles supports the dispersed engram complex hypothesis. bioRxiv 668483 [Preprint]. 12 June 2019. .doi: 10.1101/668483
100. P. W. Frankland, B. Bontempi, L. E. Talton, L. Kaczmarek, A. J. Silva, The involvement of the anterior cingulate cortex in remote contextual fear memory. *Science* **304**, 881–883 (2004). doi: 10.1126/science.1094804; pmid: 15131309
101. K. Chung *et al*., Structural and molecular interrogation of intact biological systems. *Nature* **497**, 332–337 (2013). doi: 10.1038/nature12107; pmid: 23575631
102. Y. G. Park *et al*., Protection of tissue physicochemical properties using polyfunctional crosslinkers. *Nat. Biotechnol.* (2018). pmid: 30556815
103. J. O'Keefe, J. Dostrovsky, The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* **34**, 171–175 (1971). doi: 10.1016/0006-8993(71)90358-1; pmid: 5124915
104. V. Ego-Stengel, M. A. Wilson, Disruption of ripple-associated hippocampal activity during rest impairs spatial learning in the rat. *Hippocampus* **20**, 1–10 (2010). pmid: 19816984
105. G. Girardeau, K. Benchenane, S. I. Wiener, G. Buzsáki, M. B. Zugaro, Selective suppression of hippocampal ripples impairs spatial memory. *Nat. Neurosci.* **12**, 1222–1223 (2009). doi: 10.1038/nn.2384; pmid: 19749750
106. T. J. McHugh *et al*., Dentate gyrus NMDA receptors mediate rapid pattern separation in the hippocampal network. *Science* **317**, 94–99 (2007). doi: 10.1126/science.1140263; pmid: 17556551
107. K. Z. Tanaka *et al*., The hippocampal engram maps experience but not place. *Science* **361**, 392–397 (2018). doi: 10.1126/science.aat5397; pmid: 30049878
108. M. A. Wilson, B. L. McNaughton, Reactivation of hippocampal ensemble memories during sleep. *Science* **265**, 676–679 (1994). doi: 10.1126/science.8036517; pmid: 8036517
109. S. Diekelmann, J. Born, The memory function of sleep. *Nat. Rev. Neurosci.* **11**, 114–126 (2010). doi: 10.1038/nrn2762; pmid: 20046194
110. H. R. Joo, L. M. Frank, The hippocampal sharp wave-ripple in memory retrieval for immediate use and consolidation. *Nat. Rev. Neurosci.* **19**, 744–757 (2018). doi: 10.1038/s41583-018-0077-1; pmid: 30356103
111. G. Buzsáki, Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning. *Hippocampus* **25**, 1073–1188 (2015). doi: 10.1002/hipo.22488; pmid: 26135716
112. M. F. Carr, S. P. Jadhav, L. M. Frank, Hippocampal replay in the awake state: A potential substrate for memory consolidation and retrieval. *Nat. Neurosci.* **14**, 147–153 (2011). doi: 10.1038/nn.2732; pmid: 21270783

113. H. S. Kudrimoti, C. A. Barnes, B. L. McNaughton, Reactivation of hippocampal cell assemblies: Effects of behavioral state, experience, and EEG dynamics. *J. Neurosci.* **19**, 4090–4101 (1999). doi: 10.1523/JNEUROSCI.19-10-04090.1999; pmid: 10234037
114. A. G. Siapas, M. A. Wilson, Coordinated interactions between hippocampal ripples and cortical spindles during slow-wave sleep. *Neuron* **21**, 1123–1128 (1998). doi: 10.1016/S0896-6273(00)80629-7; pmid: 9856467
115. F. Xia *et al*., Parvalbumin-positive interneurons mediate neocortical-hippocampal interactions that are necessary for memory consolidation. *eLife* **6**, e27868 (2017). doi: 10.7554/eLife.27868; pmid: 28960176
116. A. I. Abbas *et al*., Somatostatin interneurons facilitate hippocampal-prefrontal synchrony and prefrontal spatial encoding. *Neuron* **100**, 926–939.e3 (2018). doi: 10.1016/j.neuron.2018.09.029; pmid: 30318409
117. D. Dupret, J. O'Neill, B. Pleydell-Bouverie, J. Csicsvari, The reorganization and reactivation of hippocampal maps predict spatial memory performance. *Nat. Neurosci.* **13**, 995–1002 (2010). doi: 10.1038/nn.2599; pmid: 20639874
118. S. P. Jadhav, C. Kemere, P. W. German, L. M. Frank, Awake hippocampal sharp-wave ripples support spatial memory. *Science* **336**, 1454–1458 (2012). doi: 10.1126/science.1217230; pmid: 22555434
119. S. Lewis, Sleep: Ever-decreasing ripples. *Nat. Rev. Neurosci.* **19**, 184 (2018). pmid: 29515191
120. H. Norimoto *et al*., Hippocampal ripples down-regulate synapses. *Science* **359**, 1524–1527 (2018). doi: 10.1126/science.aao0702; pmid: 29439023
121. S. A. Josselyn, P. W. Frankland, Memory allocation: Mechanisms and function. *Annu. Rev. Neurosci.* **41**, 389–413 (2018). doi: 10.1146/annurev-neuro-080317-061956; pmid: 29709212
122. J. H. Han *et al*., Neuronal competition and selection during memory formation. *Science* **316**, 457–460 (2007). doi: 10.1126/science.1139438; pmid: 17446403
123. J. H. Han *et al*., Increasing CREB in the auditory thalamus enhances memory and generalization of auditory conditioned fear. *Learn. Mem.* **15**, 443–453 (2008). doi: 10.1101/lm.993608; pmid: 18519545
124. S. A. Josselyn, Continuing the search for the engram: Examining the mechanism of fear memories. *J. Psychiatry Neurosci.* **35**, 221–228 (2010). doi: 10.1503/jpn.100015; pmid: 20569648
125. A. J. Silva, Y. Zhou, T. Rogerson, J. Shobe, J. Balaji, Molecular and cellular approaches to memory allocation in neural circuits. *Science* **326**, 391–395 (2009). doi: 10.1126/science.1174519; pmid: 19833959
126. M. R. Matos *et al*., Memory strength gates the involvement of a CREB-dependent cortical fear engram in remote memory. *Nat. Commun.* **10**, 2315 (2019). doi: 10.1038/s41467-019-10266-1; pmid: 31127098
127. Y. Sano *et al*., CREB regulates memory allocation in the insular cortex. *Curr. Biol.* **24**, 2833–2837 (2014). doi: 10.1016/j.cub.2014.10.018; pmid: 25454591
128. R. Czajkowski *et al*., Encoding and storage of spatial information in the retrosplenial cortex. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8661–8666 (2014). doi: 10.1073/pnas.1313222111; pmid: 24912150
129. G. B. Choi *et al*., Driving opposing behaviors with ensembles of piriform neurons. *Cell* **146**, 1004–1015 (2011). doi: 10.1016/j.cell.2011.07.041; pmid: 21925321
130. D. Kim, D. Paré, S. S. Nair, Assignment of model amygdala neurons to the fear memory trace depends on competitive synaptic interactions. *J. Neurosci.* **33**, 14354–14358 (2013). doi: 10.1523/JNEUROSCI.2430-13.2013; pmid: 24005288
131. D. Kim, D. Paré, S. S. Nair, Mechanisms contributing to the induction and storage of Pavlovian fear memories in the lateral amygdala. *Learn. Mem.* **20**, 421–430 (2013). doi: 10.1101/lm.030262.113; pmid: 23864645
132. D. Kim, P. Samarth, F. Feng, D. Pare, S. S. Nair, Synaptic competition in the lateral amygdala and the stimulus specificity of conditioned fear: A biophysical modeling study. *Brain Struct. Funct.* **221**, 2163–2182 (2016). doi: 10.1007/s00429-015-1037-4; pmid: 25859631
133. D. L. Alkon, Changes of membrane currents during learning. *J. Exp. Biol.* **112**, 95–112 (1984). pmid: 6150967
134. D. L. Alkon, I. Lederhendler, J. J. Shoukimas, Primary changes of membrane currents during retention of associative learning. *Science* **215**, 693–695 (1982). doi: 10.1126/science.7058334; pmid: 7058334
135. K. P. Scholz, J. H. Byrne, Long-term sensitization in *Aplysia*: Biophysical correlates in tail sensory neurons. *Science* **235**, 685–687 (1987). doi: 10.1126/science.2433766; pmid: 2433766
136. M. A. Wilson, B. L. McNaughton, Dynamics of the hippocampal ensemble code for space. *Science* **261**, 1055–1058 (1993). doi: 10.1126/science.8351520 pmid: 8351520
137. J. D. Cohen, M. Bolstad, A. K. Lee, Experience-dependent shaping of hippocampal CA1 intracellular activity in novel and familiar environments. *eLife* **6**, e23040 (2017). doi: 10.7554/eLife.23040; pmid: 28742496
138. J. Epsztein, M. Brecht, A. K. Lee, Intracellular determinants of hippocampal CA1 place and silent cell activity in a novel environment. *Neuron* **70**, 109–120 (2011). doi: 10.1016/j.neuron.2011.03.006; pmid: 21482360
139. P. D. Rich, H. P. Liaw, A. K. Lee, Large environments reveal the statistical structure governing hippocampal representations. *Science* **345**, 814–817 (2014). doi: 10.1126/science.1255635; pmid: 25124440
140. D. Lee, B. J. Lin, A. K. Lee, Hippocampal place fields emerge upon single-cell manipulation of excitability during behavior. *Science* **337**, 849–853 (2012). doi: 10.1126/science.1221489; pmid: 22904011
141. J. P. Rickgauer, K. Deisseroth, D. W. Tank, Simultaneous cellular-resolution optical perturbation and imaging of place cell firing fields. *Nat. Neurosci.* **17**, 1816–1824 (2014). doi: 10.1038/nn.3866; pmid: 25402854
142. J.-P. Changeux, A. Danchin, Selective stabilisation of developing synapses as a mechanism for the specification of neuronal networks. *Nature* **264**, 705–712 (1976). doi: 10.1038/264705a0; pmid: 189195
143. J. Z. Young, Learning as a process of selection and amplification. *J. R. Soc. Med.* **72**, 801–814 (1979). doi: 10.1177/014107687907201103; pmid: 552442
144. P. Kanerva, *Sparse Distributed Memory* (MIT Press, 1988).
145. D. J. Morrison *et al*., Parvalbumin interneurons constrain the size of the lateral amygdala engram. *Neurobiol. Learn. Mem.* **135**, 91–99 (2016). doi: 10.1016/j.nlm.2016.07.007; pmid: 27422019
146. P. Rao-Ruiz, J. Yu, S. A. Kushner, S. A. Josselyn, Neuronal competition: Microcircuit mechanisms define the sparsity of the engram. *Curr. Opin. Neurobiol.* **54**, 163–170 (2019). doi: 10.1016/j.conb.2018.10.013; pmid: 30423499
147. T. Stefanelli, C. Bertollini, C. Lüscher, D. Muller, P. Mendez, Hippocampal somatostatin interneurons control the size of neuronal memory ensembles. *Neuron* **89**, 1074–1085 (2016). doi: 10.1016/j.neuron.2016.01.024; pmid: 26875623
148. H. C. Barron, T. P. Vogels, T. E. Behrens, M. Ramaswami, Inhibitory engrams in perception and memory. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 6666–6674 (2017). pmid: 28611219
149. H. C. Barron *et al*., Unmasking latent inhibitory connections in human cortex to reveal dormant cortical memories. *Neuron* **90**, 191–203 (2016). doi: 10.1016/j.neuron.2016.02.031; pmid: 26996082
150. G. Hennequin, E. J. Agnes, T. P. Vogels, Inhibitory plasticity: Balance, control, and codependence. *Annu. Rev. Neurosci.* **40**, 557–579 (2017). doi: 10.1146/annurev-neuro-072116-031005; pmid: 28598717
151. S. Maren, C. R. Ferrario, K. A. Corcoran, T. J. Desmond, K. A. Frey, Protein synthesis in the amygdala, but not the auditory thalamus, is required for consolidation of Pavlovian fear conditioning in rats. *Eur. J. Neurosci.* **18**, 3080–3088 (2003). doi: 10.1111/j.1460-9568.2003.03063.x; pmid: 14656303
152. G. E. Schafe, J. E. LeDoux, Memory consolidation of auditory Pavlovian fear conditioning requires protein synthesis and protein kinase A in the amygdala. *J. Neurosci.* **20**, RC96 (2000). doi: 10.1523/JNEUROSCI.20-18-j0003.2000; pmid: 10974093
153. P. J. Hernandez, T. Abel, The role of protein synthesis in memory consolidation: Progress amid decades of debate. *Neurobiol. Learn. Mem.* **89**, 293–311 (2008). doi: 10.1016/j.nlm.2007.09.010; pmid: 18053752
154. G. E. Schafe, N. V. Nadel, G. M. Sullivan, A. Harris, J. E. LeDoux, Memory consolidation for contextual and auditory fear conditioning is dependent on protein synthesis, PKA, and MAP kinase. *Learn. Mem.* **6**, 97–110 (1999). pmid: 10327235
155. S. A. Josselyn, S. Kida, A. J. Silva, Inducible repression of CREB function disrupts amygdala-dependent memory. *Neurobiol. Learn. Mem.* **82**, 159–163 (2004). doi: 10.1016/j.nlm.2004.05.008; pmid: 15341801
156. Y. Dudai, M. Eisenberg, Rites of passage of the engram: Reconsolidation and the lingering consolidation hypothesis. *Neuron* **44**, 93–100 (2004). doi: 10.1016/j.neuron.2004.09.003; pmid: 15450162
157. S. H. Wang, R. G. Morris, Hippocampal-neocortical interactions in memory formation, consolidation, and reconsolidation. *Annu. Rev. Psychol.* **61**, 49–79, C1–C4 (2010). doi: 10.1146/annurev.psych.093008.100523; pmid: 19575620
158. P. W. Frankland, B. Bontempi, The organization of recent and remote memories. *Nat. Rev. Neurosci.* **6**, 119–130 (2005). doi: 10.1038/nrn1607; pmid: 15685217
159. B. J. Wiltgen, R. A. M. Brown, L. E. Talton, A. J. Silva, New circuits for old memories: The role of the neocortex in consolidation. *Neuron* **44**, 101–108 (2004). doi: 10.1016/j.neuron.2004.09.015; pmid: 15450163
160. D. S. Roy, S. Muralidhar, L. M. Smith, S. Tonegawa, Silent memory engrams as the basis for retrograde amnesia. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E9972–E9979 (2017). doi: 10.1073/pnas.1714248114; pmid: 29078397
161. S. Nabavi *et al*., Engineering a memory with LTD and LTP. *Nature* **511**, 348–352 (2014). doi: 10.1038/nature13294; pmid: 24896183
162. M. H. Monfils, G. C. Teskey, Induction of long-term depression is associated with decreased dendritic length and spine density in layers III and V of sensorimotor neocortex. *Synapse* **53**, 114–121 (2004). doi: 10.1002/syn.20039; pmid: 15170823
163. Q. Zhou, K. J. Homma, M. M. Poo, Shrinkage of dendritic spines associated with long-term depression of hippocampal synapses. *Neuron* **44**, 749–757 (2004). doi: 10.1016/j.neuron.2004.11.011; pmid: 15572107
164. J. N. Bourne, K. M. Harris, Balancing structure and function at hippocampal dendritic spines. *Annu. Rev. Neurosci.* **31**, 47–67 (2008). doi: 10.1146/annurev.neuro.31.060407.125646; pmid: 18284372
165. M. F. Bear, R. C. Malenka, Synaptic plasticity: LTP and LTD. *Curr. Opin. Neurobiol.* **4**, 389–399 (1994). doi: 10.1016/0959-4388(94)90101-5; pmid: 7919934
166. D. S. Roy *et al*., Memory retrieval by activating engram cells in mouse models of early Alzheimer's disease. *Nature* **531**, 508–512 (2016). doi: 10.1038/nature17172; pmid: 26982728
167. J. N. Perusini *et al*., Optogenetic stimulation of dentate gyrus engrams restores memory in Alzheimer's disease mice. *Hippocampus* **27**, 1110–1122 (2017). doi: 10.1002/hipo.22756; pmid: 28667669
168. M. El Haj, V. Postal, P. Allain, Music enhances autobiographical memory in mild Alzheimer's disease. *Educ. Gerontol.* **38**, 30–41 (2012). doi: 10.1080/03601277.2010.515897
169. A. Herlitz, R. Adolfsson, L. Bäckman, L. G. Nilsson, Cue utilization following different forms of encoding in mildly, moderately, and severely demented patients with Alzheimer's disease. *Brain Cogn.* **15**, 119–130 (1991). doi: 10.1016/0278-2626(91)90020-9; pmid: 2009170
170. E. K. Warrington, L. Weiskrantz, Amnesic syndrome: Consolidation or retrieval? *Nature* **228**, 628–630 (1970). doi: 10.1038/228628a0; pmid: 4990853
171. R. R. Miller, L. D. Matzel, Retrieval failure versus memory loss in experimental amnesia: Definitions and processes. *Learn. Mem.* **13**, 491–497 (2006). doi: 10.1101/lm.241006; pmid: 17015845
172. O. Hardt, S. H. Wang, K. Nader, Storage or retrieval deficit: The yin and yang of amnesia. *Learn. Mem.* **16**, 224–230 (2009). doi: 10.1101/lm.1267409; pmid: 19304892
173. L. R. Squire, Lost forever or temporarily misplaced? The long debate about the nature of memory impairment. *Learn. Mem.* **13**, 522–529 (2006). doi: 10.1101/lm.310306; pmid: 17015849
174. E. Tulving, Z. Pearlstone, Availability versus accessibility of information in memory for words. *J. Verbal Learn. Verbal Behav.* **5**, 381–391 (1966). doi: 10.1016/S0022-5371(66)80048-8
175. P. W. Frankland, S. A. Josselyn, S. Köhler, The neurobiological foundation of memory retrieval. *Nat. Neurosci.* **22**, 1576–1585 (2019). doi: 10.1038/s41593-019-0493-1; pmid: 31551594
176. J. H. Kogan, P. W. Franklandand, A. J. Silva, Long-term memory underlying hippocampus-dependent social recognition in mice. *Hippocampus* **10**, 47–56 (2000). doi: 10.1002/(SICI)1098-1063(2000)10:1<47::AID-HIPO5>3.0.CO;2-6; pmid: 10706216
177. F. L. Hitti, S. A. Siegelbaum, The hippocampal CA2 region is essential for social memory. *Nature* **508**, 88–92 (2014). doi: 10.1038/nature13028; pmid: 24572357
178. T. Okuyama, T. Kitamura, D. S. Roy, S. Itohara, S. Tonegawa, Ventral CA1 neurons store social memory. *Science* **353**, 1536–1541 (2016). doi: 10.1126/science.aaf7003; pmid: 27708103
179. J. Kim *et al*., Amygdala depotentiation and fear extinction. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 20955–20960 (2007). doi: 10.1073/pnas.0710548105; pmid: 18165656
180. I. Hong *et al*., Extinction of cued fear memory involves a distinct form of depotentiation at cortical input synapses

onto the lateral amygdala. *Eur. J. Neurosci.* **30**, 2089–2099 (2009). doi: 10.1111/j.1460-9568.2009.07004.x pmid: 20128847
181. C.-H. Lin, C.-C. Lee, P.-W. Gean, Involvement of a calcineurin cascade in amygdala depotentiation and quenching of fear memory. *Mol. Pharmacol.* **63**, 44–52 (2003). doi: 10.1124/mol.63.1.44; pmid: 12488535
182. T. Yoshii, H. Hosokawa, N. Matsuo, Pharmacogenetic reactivation of the original engram evokes an extinguished fear memory. *Neuropharmacology* **113**, 1–9 (2017). doi: 10.1016/j.neuropharm.2016.09.012; pmid: 27639988
183. R. L. Clem, D. Schiller, New learning and unlearning: Strangers or accomplices in threat memory attenuation? *Trends Neurosci.* **39**, 340–351 (2016). doi: 10.1016/j.tins.2016.03.003; pmid: 27079843
184. N. V. Luchkina, V. Y. Bolshakov, Mechanisms of fear learning and extinction: Synaptic plasticity-fear memory connection. *Psychopharmacology (Berl.)* **236**, 163–182 (2019). doi: 10.1007/s00213-018-5104-4; pmid: 30415278
185. C. A. Orsini, S. Maren, Neural and cellular mechanisms of fear and extinction memory formation. *Neurosci. Biobehav. Rev.* **36**, 1773–1802 (2012). doi: 10.1016/j.neubiorev.2011.12.014; pmid: 22230704
186. K. M. Myers, M. Davis, Mechanisms of fear extinction. *Mol. Psychiatry* **12**, 120–150 (2007). doi: 10.1038/sj.mp.4001939; pmid: 17160066
187. M. E. Bouton, R. F. Westbrook, K. A. Corcoran, S. Maren, Contextual and temporal modulation of extinction: Behavioral and biological mechanisms. *Biol. Psychiatry* **60**, 352–360 (2006). doi: 10.1016/j.biopsych.2005.12.015; pmid: 16616731
188. J. Ji, S. Maren, Hippocampal involvement in contextual modulation of fear extinction. *Hippocampus* **17**, 749–758 (2007). doi: 10.1002/hipo.20331; pmid: 17604353
189. M. E. Bouton, Context and behavioral processes in extinction. *Learn. Mem.* **11**, 485–494 (2004). doi: 10.1101/lm.78804; pmid: 15466298
190. G. J. Quirk, D. Mueller, Neural mechanisms of extinction learning and retrieval. *Neuropsychopharmacology* **33**, 56–72 (2008). doi: 10.1038/sj.npp.1301555; pmid: 17882236
191. M. E. Bouton, J. B. Nelson, Context-specificity of target versus feature inhibition in a feature-negative discrimination. *J. Exp. Psychol. Anim. Behav. Process.* **20**, 51–65 (1994). doi: 10.1037/0097-7403.20.1.51; pmid: 8308493
192. R. A. Rescorla, C. D. Heth, Reinstatement of fear to an extinguished conditioned stimulus. *J. Exp. Psychol. Anim. Behav. Process.* **1**, 88–96 (1975). doi: 10.1037/0097-7403.1.1.88; pmid: 1151290
193. M. H. Monfils, K. K. Cowansage, E. Klann, J. E. LeDoux, Extinction-reconsolidation boundaries: Key to persistent attenuation of fear memories. *Science* **324**, 951–955 (2009). doi: 10.1126/science.1167975; pmid: 19342552
194. D. Schiller *et al*., Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature* **463**, 49–53 (2010). doi: 10.1038/nature08637; pmid: 20010606
195. X. Zhang, J. Kim, S. Tonegawa, Amygdala reward neurons form and store fear extinction memory. *Neuron* 10.1016/j.neuron.2019.12.025 (2020).
196. J. J. Kim, M. S. Fanselow, Modality-specific retrograde amnesia of fear. *Science* **256**, 675–677 (1992). doi: 10.1126/science.1585183; pmid: 1585183
197. S. Maren, G. Aharonov, M. S. Fanselow, Neurotoxic lesions of the dorsal hippocampus and Pavlovian fear conditioning in rats. *Behav. Brain Res.* **88**, 261–274 (1997). doi: 10.1016/S0166-4328(97)00088-0; pmid: 9404635
198. K. K. Tayler, K. Z. Tanaka, L. G. Reijmers, B. J. Wiltgen, Reactivation of neural ensembles during the retrieval of recent and remote memory. *Curr. Biol.* **23**, 99–106 (2013). doi: 10.1016/j.cub.2012.11.019
199. S. Tonegawa, M. D. Morrissey, T. Kitamura, The role of engram cells in the systems consolidation of memory. *Nat. Rev. Neurosci.* **19**, 485–498 (2018). doi: 10.1038/s41583-018-0031-2; pmid: 29970909
200. B. J. Wiltgen, A. J. Silva, Memory for context becomes less specific with time. *Learn. Mem.* **14**, 313–317 (2007). doi: 10.1101/lm.430907; pmid: 17522020
201. B. J. Wiltgen *et al*., The hippocampus plays a selective role in the retrieval of detailed contextual memories. *Curr. Biol.* **20**, 1336–1344 (2010). doi: 10.1016/j.cub.2010.06.068 pmid: 20637623
202. S. H. Wang, C. M. Teixeira, A. L. Wheeler, P. W. Frankland, The precision of remote context memories does not require the hippocampus. *Nat. Neurosci.* **12**, 253–255 (2009). doi: 10.1038/nn.2263; pmid: 19182794
203. G. Winocur, M. Moscovitch, Memory transformation and systems consolidation. *J. Int. Neuropsychol. Soc.* **17**, 766–780 (2011). doi: 10.1017/S1355617711000683; pmid: 21729403
204. M. Moscovitch, R. Cabeza, G. Winocur, L. Nadel, Episodic memory and beyond: The hippocampus and neocortex in transformation. *Annu. Rev. Psychol.* **67**, 105–134 (2016). doi: 10.1146/annurev-psych-113011-143733; pmid: 26726963
205. N. Guo *et al*., Dentate granule cell recruitment of feedforward inhibition governs engram maintenance and remote memory generalization. *Nat. Med.* **24**, 438–449 (2018). doi: 10.1038/nm.4491; pmid: 29529016
206. M. Shehata *et al*., Autophagy enhances memory erasure through synaptic destabilization. *J. Neurosci.* **38**, 3809–3822 (2018). doi: 10.1523/JNEUROSCI.3505-17.2018; pmid: 29555855
207. C. B. Kirwan, C. E. Stark, Overcoming interference: An fMRI investigation of pattern separation in the medial temporal lobe. *Learn. Mem.* **14**, 625–633 (2007). doi: 10.1101/lm.663507; pmid: 17848502
208. J. K. Leutgeb, S. Leutgeb, M. B. Moser, E. I. Moser, Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science* **315**, 961–966 (2007). doi: 10.1126/science.1135801; pmid: 17303747
209. K. A. Norman, R. C. O'Reilly, Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychol. Rev.* **110**, 611–646 (2003). doi: 10.1037/0033-295X.110.4.611; pmid: 14599236
210. A. Gilboa, H. Marlatte, Neurobiology of schemas and schema-mediated memory. *Trends Cogn. Sci.* **21**, 618–631 (2017). doi: 10.1016/j.tics.2017.04.013; pmid: 28551107
211. S. McKenzie, H. Eichenbaum, Consolidation and reconsolidation: Two lives of memories? *Neuron* **71**, 224–233 (2011). doi: 10.1016/j.neuron.2011.06.037; pmid: 21791282
212. M. L. Schlichting, A. R. Preston, Memory integration: Neural mechanisms and implications for behavior. *Curr. Opin. Behav. Sci.* **1**, 1–8 (2015). doi: 10.1016/j.cobeha.2014.07.005; pmid: 25750931
213. D. Tse *et al*., Schemas and memory consolidation. *Science* **316**, 76–82 (2007). doi: 10.1126/science.1135935; pmid: 17412951
214. D. Zeithamova, A. R. Preston, Temporal proximity promotes integration of overlapping events. *J. Cogn. Neurosci.* **29**, 1311–1323 (2017). doi: 10.1162/jocn_a_01116; pmid: 28253077
215. A. J. Rashid *et al*., Competition between engrams influences fear memory formation and recall. *Science* **353**, 383–387 (2016). doi: 10.1126/science.aaf0594; pmid: 27463673
216. D. J. Cai *et al*., A shared neural ensemble links distinct contextual memories encoded close in time. *Nature* **534**, 115–118 (2016). doi: 10.1038/nature17955; pmid: 27251287
217. M. Sehgal *et al*., Memory allocation mechanisms underlie memory linking across time. *Neurobiol. Learn. Mem.* **153**, 21–25 (2018). doi: 10.1016/j.nlm.2018.02.021; pmid: 29496645
218. P. Rao-Ruiz *et al*., Engram-specific transcriptome profiling of contextual memory consolidation. *Nat. Commun.* **10**, 2232 (2019). doi: 10.1038/s41467-019-09960-x; pmid: 31110186
219. M. Pignatelli *et al*., Engram cell excitability state determines the efficacy of memory retrieval. *Neuron* **101**, 274–284.e5 (2019). doi: 10.1016/j.neuron.2018.11.029; pmid: 30551997
220. J. Yokose *et al*., Overlapping memory trace indispensable for linking, but not recalling, individual memories. *Science* **355**, 398–403 (2017). doi: 10.1126/science.aal2690; pmid: 28126819
221. B. J. Levy, A. D. Wagner, Measuring memory reactivation with functional MRI: Implications for psychological theory. *Perspect. Psychol. Sci.* **8**, 72–78 (2013). doi: 10.1177/1745691612469031; pmid: 25484909
222. M. L. Mack, B. C. Love, A. R. Preston, Building concepts one episode at a time: The hippocampus and concept formation. *Neurosci. Lett.* **680**, 31–38 (2018). doi: 10.1016/j.neulet.2017.07.061; pmid: 28801273
223. M. L. Schlichting, A. R. Preston, Memory reactivation during rest supports upcoming learning of related content. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15845–15850 (2014). doi: 10.1073/pnas.1404396111; pmid: 25331890
224. D. Zeithamova, M. L. Schlichting, A. R. Preston, The hippocampus and inferential reasoning: Building memories to navigate future decisions. *Front. Hum. Neurosci.* **6**, 70 (2012). doi: 10.3389/fnhum.2012.00070; pmid: 22470333
225. M. S. Gazzaniga, *Conversations in the Cognitive Neurosciences* (MIT Press, 1997).
226. H. B. Barlow, in *The Cognitive Neurosciences*, M. S. Gazzaniga, Ed. (MIT Press, 1995), pp. 415–435.
227. C. G. Gross, Genealogy of the "grandmother cell". *Neuroscientist* **8**, 512–518 (2002). doi: 10.1177/107385802237175; pmid: 12374433
228. R. Yuste, From the neuron doctrine to neural networks. *Nat. Rev. Neurosci.* **16**, 487–497 (2015). doi: 10.1038/nrn3962; pmid: 26152865
229. H. Eichenbaum, Barlow versus Hebb: When is it time to abandon the notion of feature detectors and adopt the cell assembly as the unit of cognition? *Neurosci. Lett.* **680**, 88–93 (2018). doi: 10.1016/j.neulet.2017.04.006; pmid: 28389238
230. C. A. Denny, E. Lebois, S. Ramirez, From engrams to pathologies of the brain. *Front. Neural Circuits* **11**, 23 (2017). doi: 10.3389/fncir.2017.00023; pmid: 28439228
231. E. C. Tolman, C. H. Honzik, Introduction and removal of reward, and maze performance in rats. *Univ. Calif. Publ. Psychol.* **4**, 257–275 (1930).
232. G. T. Philips, E. I. Tzvetkova, S. Marinesco, T. J. Carew, Latent memory for sensitization in *Aplysia*. *Learn. Mem.* **13**, 224–229 (2006). doi: 10.1101/lm.111506; pmid: 16585798
233. R. E. Lubow, A. U. Moore, Latent inhibition: The effect of nonreinforced pre-exposure to the conditional stimulus. *J. Comp. Physiol. Psychol.* **52**, 415–419 (1959). doi: 10.1037/h0046700; pmid: 14418647
234. D. J. Lewis, Psychobiology of active and inactive memory. *Psychol. Bull.* **86**, 1054–1083 (1979). doi: 10.1037/0033-2909.86.5.1054; pmid: 386401
235. K. Nader, G. E. Schafe, J. E. Le Doux, Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature* **406**, 722–726 (2000). doi: 10.1038/35021052; pmid: 10963596
236. J. Przybyslawski, S. J. Sara, Reconsolidation of memory after its reactivation. *Behav. Brain Res.* **84**, 241–246 (1997). doi: 10.1016/S0166-4328(96)00153-2; pmid: 9079788
237. S. Kida *et al*., CREB required for the stability of new and reactivated fear memories. *Nat. Neurosci.* **5**, 348–355 (2002). doi: 10.1038/nn819; pmid: 11889468
238. S. J. Sara, Retrieval and reconsolidation: Toward a neurobiology of remembering. *Learn. Mem.* **7**, 73–84 (2000). doi: 10.1101/lm.7.2.73; pmid: 10753974
239. R. G. Morris *et al*., Memory reconsolidation: Sensitivity of spatial memory to inhibition of protein synthesis in dorsal hippocampus during encoding and retrieval. *Neuron* **50**, 479–489 (2006). doi: 10.1016/j.neuron.2006.04.012; pmid: 16675401
240. J. L. Lee, Memory reconsolidation mediates the updating of hippocampal memory content. *Front. Behav. Neurosci.* **4**, 168 (2010). doi: 10.3389/fnbeh.2010.00168; pmid: 21120142
241. M. Eisenberg, T. Kobilo, D. E. Berman, Y. Dudai, Stability of retrieved memory: Inverse correlation with trace dominance. *Science* **301**, 1102–1104 (2003). doi: 10.1126/science.1086881; pmid: 12934010
242. M. C. Anderson, R. A. Bjork, E. L. Bjork, Remembering can cause forgetting: Retrieval dynamics in long-term memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **20**, 1063–1087 (1994). doi: 10.1037/0278-7393.20.5.1063; pmid: 7931095
243. B. A. Richards, P. W. Frankland, The conjunctive trace. *Hippocampus* **23**, 207–212 (2013). doi: 10.1002/hipo.22089; pmid: 23389924
244. A. Rubin, N. Geva, L. Sheintuch, Y. Ziv, Hippocampal ensemble dynamics timestamp events in long-term memory. *eLife* **4**, e12247 (2015). doi: 10.7554/eLife.12247; pmid: 26682652
245. S. Brodt *et al*., Fast track to the neocortex: A memory engram in the posterior parietal cortex. *Science* **362**, 1045–1048 (2018). doi: 10.1126/science.aau2528; pmid: 30498125
246. K. Nader, Memory traces unbound. *Trends Neurosci.* **26**, 65–72 (2003). doi: 10.1016/S0166-2236(02)00042-5; pmid: 12536129
247. Y. Dudai, The restless engram: Consolidations never end. *Annu. Rev. Neurosci.* **35**, 227–247 (2012). doi: 10.1146/annurev-neuro-062111-150500; pmid: 22443508
248. D. Schiller, E. A. Phelps, Does reconsolidation occur in humans? *Front. Behav. Neurosci.* **5**, 24 (2011). doi: 10.3389/fnbeh.2011.00024; pmid: 21629821
249. M. Kindt, M. Soeter, B. Vervliet, Beyond extinction: Erasing human fear responses and preventing the return of fear. *Nat. Neurosci.* **12**, 256–258 (2009). doi: 10.1038/nn.2271; pmid: 19219038

ACKNOWLEDGMENTS

We thank our many colleagues for interesting conversations that shaped this review. In particular, we would like to acknowledge the contributions of Y. Dudai, P. Frankland, S. Köhler, M. Pignatelli, and S. Waddell, as well as J. Lau (for figure preparation) and D. Roy and J. Yu (for a sorted publication list); and the members of the Josselyn, Tonegawa, and Frankland labs for helpful discussions. **Funding:** Supported by the Canadian Institute of Health Research (CIHR, FDN-388455), the Natural Science and Engineering Research Council (NSERC) Discovery Grant, the Canadian Institute for Advanced Studies (CiFAR) Grant, and the NIH (NIMH, 1 R01 MH119421-01) (to S.A.J); and by RIKEN's Center for Brain Science, Howard Hughes Medical Institute (HHMI), and JPB Foundation (to S.T.). **Competing interests:** The authors declare no competing interests.

10.1126/science.aaw4325

RESEARCH ARTICLE

NANOPARTICLES

Oriented attachment induces fivefold twins by forming and decomposing high-energy grain boundaries

Miao Song[1]*, Gang Zhou[2]*, Ning Lu[3]*, Jaewon Lee[1,4], Elias Nakouzi[1], Hao Wang[2], Dongsheng Li[1]†

Natural and synthetic nanoparticles composed of fivefold twinned crystal domains have distinct properties. The formation mechanism of these fivefold twinned nanoparticles is poorly understood. We used in situ high-resolution transmission electron microscopy combined with molecular dynamics simulations to demonstrate that fivefold twinning occurs through repeated oriented attachment of ~3-nanometer gold, platinum, and palladium nanoparticles. We discovered two different mechanisms for forming fivefold twinned nanoparticles that are driven by the accumulation and elimination of strain. This was accompanied by decomposition of grain boundaries and the formation of a special class of twins with a net strain of zero. These observations allowed us to develop a quantitative picture of the twinning process. The mechanisms provide guidance for controlling twin structures and morphologies across a wide range of materials.

Twinning in materials occurs when two crystals that share the same crystal lattice plane intergrow through certain symmetry operations. Crystallographic twins are widespread, being found in a wide range of materials that include minerals (such as rutile and wurtzite) (*1*, *2*), metals [such as copper (Cu), silver (Ag), and gold (Au)] (*3*, *4*), and ceramics [barium titanate ($BaTiO_3$)] (*5*). Twinning leads to a variety of structures and morphologies that affect physical and chemical properties. For example, the stress of fivefold twins (5-FTs) substantially increases the Young's modulus of nanowires (*6*), whereas multitwinned Cu nanowires exhibit excellent methane selectivity during reduction of carbon dioxide (*7*). Multiply twinned structures have attracted substantial attention for applications in crystal growth (*8*), mechanical engineering (*9*), biomedical diagnosis (*10*), optics (*11*), and catalysis (*12*).

Understanding the properties of these twinned materials required determining the formation mechanisms that enables control of the growth and the final shape. Fivefold twins are a common multiply twinned structure that was discovered nearly 200 years ago (*13*, *14*), and the formation mechanisms and potential applications have been widely studied (*15*, *16*). However, the underlying formation mechanisms are unclear and subject to debate because of the challenges of making direct observations of the formation process at the atomic scale. Proposed formation mechanisms include (i) direct nucleation through atom-by-atom addition on the basis of theoretical simulations (*17*), (ii) the successive twining and growth of tetrahedral units (*4*, *18*, *19*), and (iii) partial dislocation slipping (*20*). Experimental studies were mainly based on morphology evolution, in contrast to the decahedral symmetry of 5-FTs (*4*, *18*, *19*). Furthermore, in these experimental studies, the particles had already undergone twinning by the time they were observed (*18*, *20*). Oriented attachment (OA) of particles, which is a candidate pathway for 5-FT formation, was observed directly during the formation of twin interfaces (*21*). We primarily focused on face-centered cubic Au nanoparticles (NPs) as a model system with which to investigate the twinning process at the atomic scale. In order to observe the onset of twinning, we chose small (~3 nm) Au NPs because of the size effect on 5-FTs, which are stable with the size of 3 to 14 nm owing to thermodynamics (*22*–*24*). Twinning exists in Au NPs of ≥3 nm in our experiment (fig. S1, B and E) (*25*).

Multiply twinned structures formed through OA

We drop-cast spherical Au NPs (~3 nm) (fig. S1) embedded in an organic matrix of 1-dodecanethiol (fig. S2) on a transmission electron microscope (TEM) grid. The organics decomposed under electron-beam (e-beam) irradiation and the surface of the Au NPs self-adjusted to decrease the surface energy

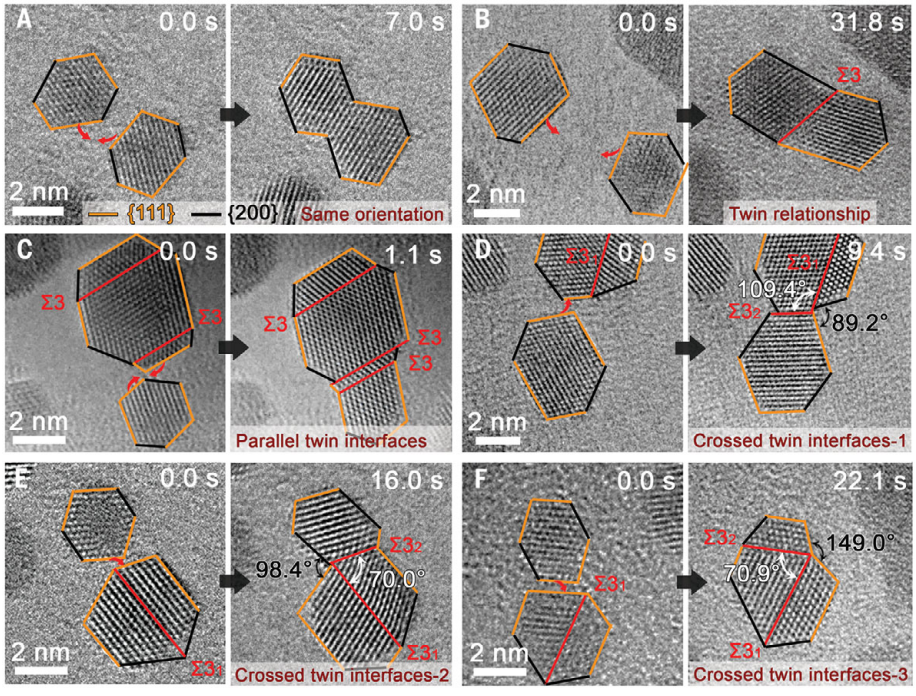


Fig. 1. Examples of resulting crystal structures after OA processes of Au NPs. (**A**) Formation of a single crystal. (**B**) Formation of a twin structure with a twin boundary. (**C** to **F**) Formation of crystals with multiple twin interfaces, including (C) parallel twin interfaces, (D) crossed twin interfaces with an angle of ~109° between Σ3s and a concave surface angle of ~89°, (E) crossed twin interfaces with an angle of ~70° between Σ3s and a concave surface angle of ~98°, and (F) crossed twin interfaces with an angle of ~71° between Σ3s and a concave surface angle of ~149°.

[1]Physical and Computational Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA 99352, USA. [2]Division of Titanium Alloys, Institute of Metal Research, Chinese Academy of Sciences, Shenyang 110016, China. [3]Department of Materials Science and Engineering, University of Michigan, Ann Arbor, MI 48109, USA. [4]Department of Biomedical, Biological and Chemical Engineering, University of Missouri, Columbia, MO 65211, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: dongsheng.li2@pnnl.gov

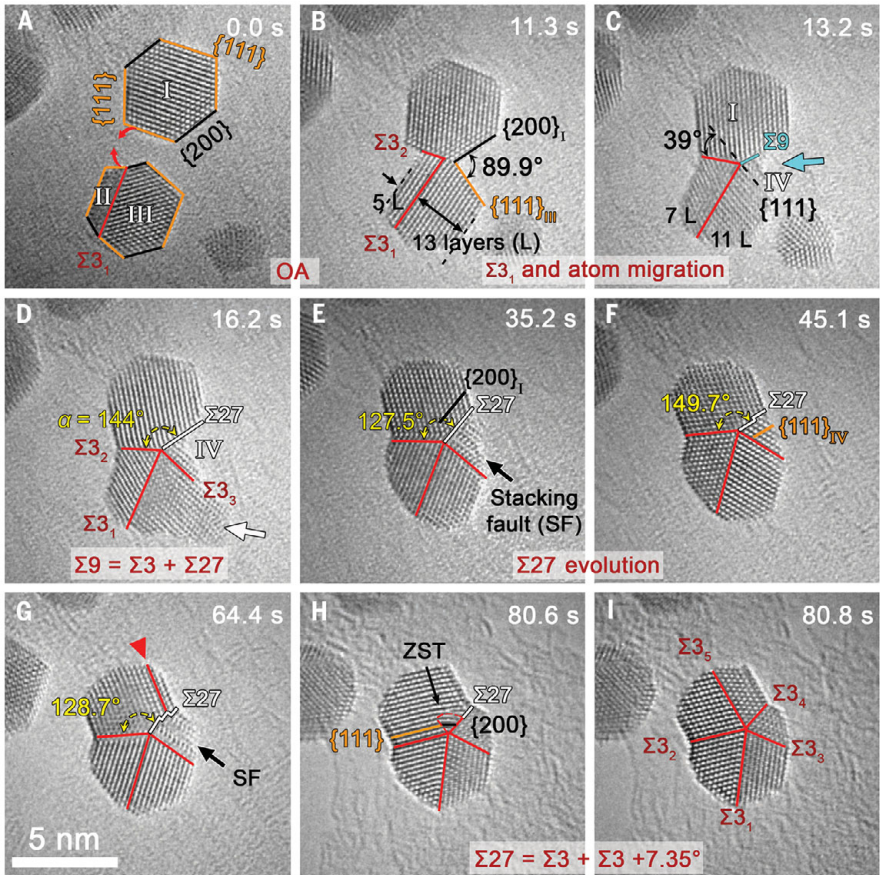


Fig. 2. 5-FT formation mechanism 1 through OA, atomic surface diffusion, and after nucleation and growth of ZST. (**A** and **B**) Formation of $\Sigma3_2$ and ~90° concave surface after the OA process. (**C**) Migration of atoms to the concave surface region (blue arrow) with the formation of the Σ9 GB and $\Sigma3_2$ migration of two atomic layers toward region III from II. (**D** to **G**) Oscillation between $\Sigma27\text{-}(200)_I$ and $\Sigma27\text{-}(111)_{IV}$ and corresponding α. (**H**) Nucleation of ZST on Σ27 near the twin pole. (**I**) Formation of 5-FT. The red arrowhead in (G) denotes one layer of twin interface through partial dislocation slipping.

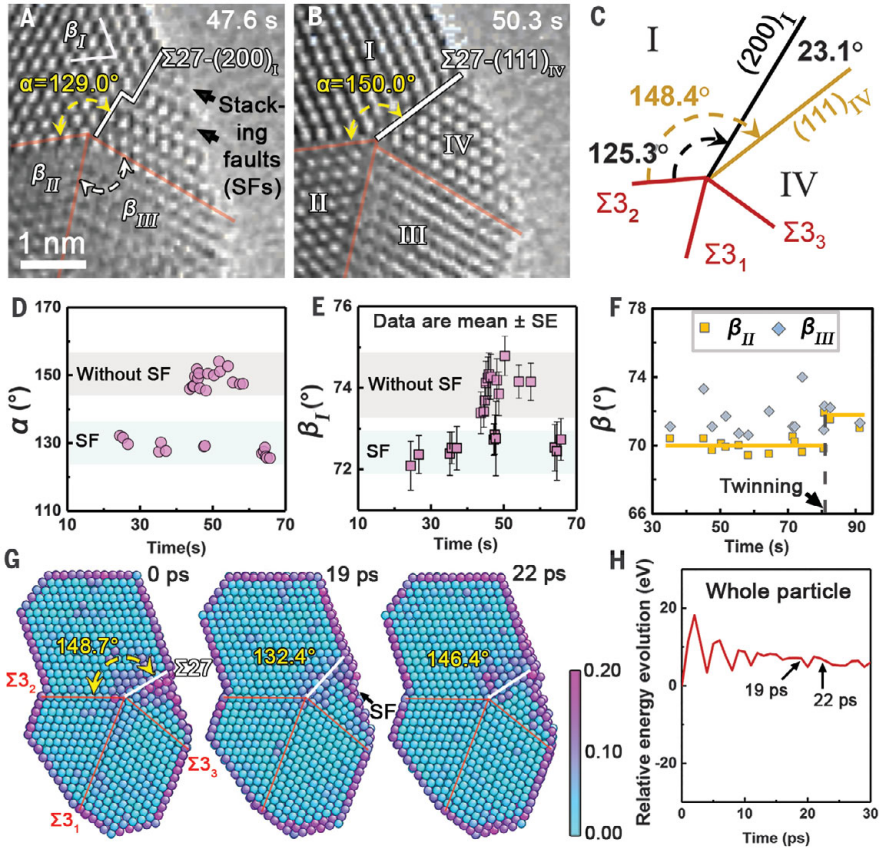


Fig. 3. Σ27 oscillatory evolution process before the 5-FT formation. (**A** and **B**) Enlarged high-resolution TEM images of the NP in Fig. 2 showing α oscillation. (**C**) Theoretical angle misfit between $(200)_I$ and $(111)_{IV}$. (**D** to **F**) Variation of α, β_I, β_{II}, and β_{III} with time. Approximately 20 to 80 measurements were taken for each data point in Fig. 3E. SE, standard error. (**G** and **H**) Atomic structure and energy profile of the whole NP during the oscillatory evolution process of Σ27 based on MD simulation, respectively.

(*26*), presenting a hexagonal two-dimensional (2D) projection composed of four {111} and two {200} planes (Fig. 1A). During this process, the Au NPs moved and aggregated. The decomposition of the surfactant led us to make the assumption that the type does not affect the mechanisms for this system (*26*). To understand the relation between the two grains, we used the degree of fit (Σ) to define the grain boundary (GB) based on coincident site lattice theory, where Σ is the ratio of the total number of sites to coincidence sites (*27*). During aggregation, NPs specifically underwent OA along the {111} planes, forming either interface-free single crystals (Fig. 1A) or Σ3 twin structures (Fig. 1B), similar to Au and iron oxide particle OA processes previously reported in liquid phases (*21*, *28*). In addition, multiple twin interfaces often formed. For example, repeated OA processes induced both parallel (Fig. 1C) and three types of crossed-twin interfaces (Fig. 1, D to F). A second twin boundary $\Sigma3_2$ formed when a twinned crystal (with the first $\Sigma3_1$) underwent OA with another single crystal, resulting in two types of concave surfaces with angles of ~94° ({111}+{200} surfaces) (Figs. 1, D and E, and 2, A and B) and ~150° (two {111} surfaces) (Fig. 1F). The angle between two Σ3s was either ~109° (Fig. 1D) or ~70° (Fig. 1, E and F). In addition, we often observed twinning and detwinning processes through the slipping of atoms in {111} planes by $\frac{1}{6}$[112]—that is, partial dislocation slipping with a Burgers vector (*b*) of $\frac{1}{6}$[112] (fig. S3).

Mechanism 1: Formation of 5-FT by forming and decomposing high-energy grain boundaries

The angle of ~70° between two Σ3s (Fig. 1, E and F) provided the possibility of 5-FT formation, in which the theoretical angle between Σ3s is 70.53°. The two types of concave surfaces with angles of ~98° (Fig. 1E) and ~150° (Fig. 1F) lead to two mechanisms for the 5-FT formation. At the 89.9° concave surface (Fig. 2B and movie S1, the same OA process as shown in Fig. 1E), adjacent atoms migrated to this high-energy surface because of the small curvature radius (Fig. 2C and movie S1) (*29*), forming a new high-energy Σ9, identified by the 39° angle between {111} planes in regions I and IV (Fig. 2C and fig. S4, A to C). The GB energy (E_{GB}) of Σ9 is ~542 mJ/m^2 (*30*), which is high compared with the ~17.5 mJ/m^2 E_{GB} of Σ3 (*31*). Because of this high E_{GB}, Σ9 subsequently decomposed into a stable Σ3 (the third $\Sigma3_3$) and a new high-energy Σ27 (E_{GB} ~ 560 mJ/m^2) (*30*), identified by the 32° angle between the {111} planes in regions I and IV (Fig. 2D and fig. S4, D to F). The Σ9 decomposition follows the dissociation theory of the coincidence site lattice [Σ(A/B) ↔ ΣA + ΣB or Σ(A × B) ↔ ΣA + ΣB] (*32*). As atoms further filled the concave surface through either particle aggregation (Fig. 2, C and D) or surface atom diffusion (fig. S5), stacking faults formed in region IV, resulting in a Σ27 between stacking faults and $\{200\}_I$ planes in region I [$\Sigma27\text{-}(200)_I$] (Fig. 2E). This unstable $\Sigma27\text{-}(200)_I$ evolved into a different Σ27 between $(111)_{IV}$ and the region I [$\Sigma27\text{-}(111)_{IV}$] with the disappearance of stacking faults (Fig. 2F and movie S1), resulting in an oscillation between them (Figs. 2, E to G, and 3). After three consecutive 64-s oscillation cycles (Fig. 2, D to H; fig. S6; and movie S1), twin lamellae—previously termed a zero-strain twin (ZST)—that exhibit localized strain, but possess strain fields that sum to zero in the surrounding crystal lattice, nucleated at the twin pole along Σ27 (Figs. 2H and 4) (*33*, *34*). The Σ27 immediately decomposed into the fourth and fifth twin boundaries ($\Sigma3_4$ and $\Sigma3_5$) within 0.2 s and formed a 5-FT (Fig. 2I), corresponding to $\Sigma27 \rightarrow \Sigma3_4 + \Sigma3_5 + 7.35°$ (Σ81 = $\Sigma3_5$ + 7.35°). The misfit angle of 7.35° between the twin planes after 5-FT formation is accommodated by lattice strain in twin units (*30*), and the lattice strain can also be relaxed through surface adjustment, such as forming a truncated (Marks) decahedron (fig. S7) (*35*).

The evolution between $\Sigma27\text{-}(200)_I$ and $\Sigma27\text{-}(111)_{IV}$ was accompanied by an oscillation of the angle α (between Σ27 and $\Sigma3_2$) (Figs. 2, E to G, and 3, A to D). The theoretical α is 125.3° and 148.4° for $\Sigma27\text{-}(200)_I$ and $\Sigma27(111)_{IV}$, respectively, which leads to a large angle misfit of 23.1° between $(200)_I$ and $(111)_{IV}$ (Fig. 3C). During this process, the angle misfit was mainly accommodated by either stacking faults in region IV or lattice deformation in region I. In a perfect Au crystal, the angle (β) between {111} planes is 70.53°. With stacking faults (α ~ 130°) (Fig. 3A), the lattice deformed slightly in region I, with β_I of 72.5°. When the stacking faults disappeared (α ~ 150°), the lattice in region I deformed further, with β_I of ~74° (Fig. 3E), to compensate the lattice mismatch. Region III was also affected during the oscillatory evolution of Σ27, with

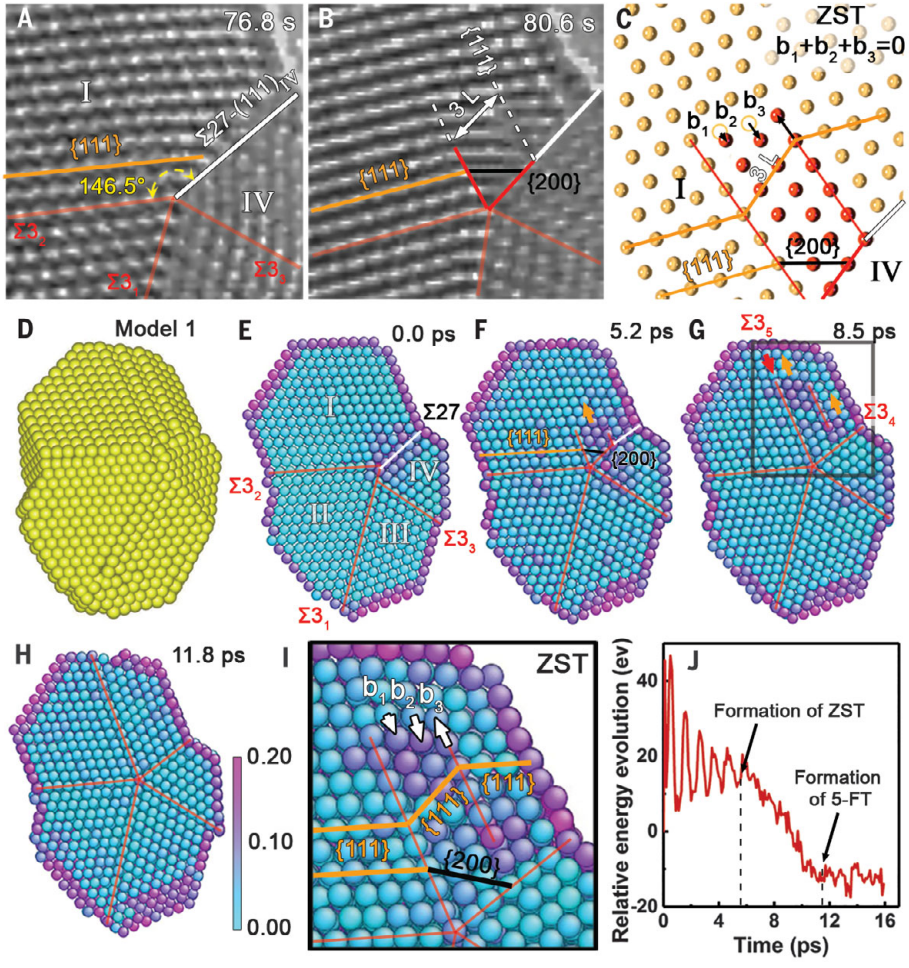


Fig. 4. Decomposition of Σ27 through ZST forming $\Sigma3_4$ and $\Sigma3_5$. (**A** and **B**) Enlarged high-resolution TEM images of twin pole regions of the NP at 76.8 and 80.6 s, respectively, showing the formation of a ZST near the twin pole. (**C**) Schematics of the ZST. (**D**) MD simulation (model 1) with {111} surfaces. (**E** to **H**) ZST nucleation on Σ27 and growth in region I at 800 K. (**I**) Enlarged black-box area in (G). (**J**) Energy profile of the NP during decomposition of Σ27.

β_{III} ranging from 71° to 74° (~72° on average) (Fig. 3F). By contrast, region II was barely affected, and β_{II} remained at ~70° (Fig. 3F). However, β_{II} increased instantly to ~72° as the 5-FT formed at 80.6 s (Fig. 3F) and remained at ~72°, sharing the angle misfit (7.35°) with the other twin units. Molecular dynamics (MD) simulations confirmed a similar oscillation (Fig. 3G), showing comparable energies between them (Fig. 3H). MD results also consistently demonstrated the relatively minor shear strain in region I with stacking faults (Fig. 3G, α ~ 132.4°) and an increase of shear strain near the Σ27 boundary without stacking faults (Fig. 3G, α ~ 146.4°).

As described above, region I was highly strained without stacking faults, with β_I of ~74°; the shear strain was concentrated near the twin pole along Σ27-$(111)_{IV}$ (Fig. 4, A and E), where at 80.6 s, three layers of {111} planes deformed into {200} planes (Fig. 4B and fig. S8). This deformation did not affect the neighboring lattice structures, indicating nucleation of a ZST. A ZST is typically formed by atoms deforming from {111} into {200} planes through partial dislocation slipping of three consecutive {111} planes (Fig. 4C) with Burgers vectors (b) of $b_1 = \frac{1}{6}[2\bar{1}\bar{1}]$, $b_2 = \frac{1}{6}[\bar{1}12]$, and $b_3 = \frac{1}{6}[\bar{1}2\bar{1}]$, resulting in a total b of zero (Fig. 4C) (*33*, *34*). We found with MD simulation that the shear strain concentration on the Σ27-$\{111\}_{IV}$ (Fig. 4E) induces nucleation of ZST (Fig. 4, D to G and I; figs. S9 and S10; and movies S2 and S3). The ZST grows laterally along the three layers of the $\{111\}_I$ planes in region I (Fig. 4, F and G, yellow arrows), resulting in strain relaxation on the unstable Σ27 (Fig. 4G) and formation of $\Sigma3_5$ in region I (Fig. 4G). The continuous nucleation and growth of ZST (movies S2 and S3) lead to the complete decomposition of Σ27-$(111)_{IV}$ into $\Sigma3_4$ and $\Sigma3_5$, forming the 5-FT (Fig. 4H) with a large energy decrease (Fig. 4J). Residual layers of $\{111\}_I$ planes (less than three layers) near the NP surface can be twinned through partial dislocation slipping (Fig. 2G). Fully decomposing Σ27 through {111} partial dislocation slipping nucleated on the surface of the NP is unfavorable because of the angle mismatch between {111} planes of regions I and IV (fig. S11), although it has been observed to form a few layer twin units. There was a substantial configuration adjustment of the NP during the Σ27 oscillatory evolution process (Fig. 2, D to H, and fig. S12), indicating surface adjustment of the NP. During this period, a nucleation attempt of the ZST at 71.0 s occurred, but it disappeared in ~0.3 s (fig. S13), indicating a failure to overcome the energy barrier to nucleate and grow.

Mechanism 2: Formation of 5-FT through partial dislocation slipping

At the resulting ~150° concave surface (Fig. 5, A to C), the surface is relatively more stable than the 94° concave surface, and the atoms seldom migrated to the concave surface area. Under e-beam irradiation, one layer of twin structures formed at the concave area through GB decomposition and partial dislocation slipping (Fig. 5, C, D, and G; and fig. S14). Twin interfaces subsequently migrated from the periphery of the particle toward the center and formed $\Sigma3_3$, $\Sigma3_4$, and $\Sigma3_5$ (Fig. 5, D and E). On the basis of our eight observations of this process, twin boundaries only migrate one to five atomic layers, resulting in the unstable asymmetric 5-FT (Fig. 5, D and E, and fig. S15). The asymmetric 5-FT underwent detwinning through partial dislocation slipping, as demonstrated by the varying number of {111} layers in each twin unit (Fig. 5, E and F), to decrease the twin interface and volume strain energies (*36*). During detwinning processes, twin boundary migration and pole splitting were observed (more examples are provided in figs. S16 and S17), similar to what we reported previously, the detwinning of Ag 5-FT particles in solution (*36*).

In addition, the asymmetric 5-FT can also grow into a stable structure through particle aggregation (OA or non-OA) (fig. S18, A to E). Similarly, a symmetrical 5-FT can evolve into an asymmetrical structure through particle aggregation (fig. S18, F to J). On the basis of 15 observations of 5-FT formation, the probability of these two twinning mechanisms occurring is almost equal (7:8) because of the random selectivity of the attached {111} surfaces during OA. We collected a number of other examples (figs. S15 and S17) of these two types of twinning processes (fig. S19).

Formation of 5-FT through OA in other metal systems

We observed the two fivefold twinning mechanisms not only in Au NPs with a low twin-fault energy (~26 mJ/m^2) (*37*) but also in other metal NP systems with high twin-fault energies, palladium (Pd) NPs (~106 mJ/m^2) (Fig. 6, A to H) (*37*) and platinum (Pt) NPs (~179 mJ/m^2) (Fig. 6, I to P) (*37*). For example, a twinned NP (Fig. 6A) underwent OA with another NP (Fig. 6B), inducing the second $\Sigma3_2$ and a ~85° concave surface; after atomic surface diffusion, a high-energy Σ9 GB formed (Fig. 6, A and B). Subsequently, Σ9 decomposed into another high-energy GB Σ27 and the third $\Sigma3_3$ (Fig. 6, B and C). Last, Σ27 evolved into

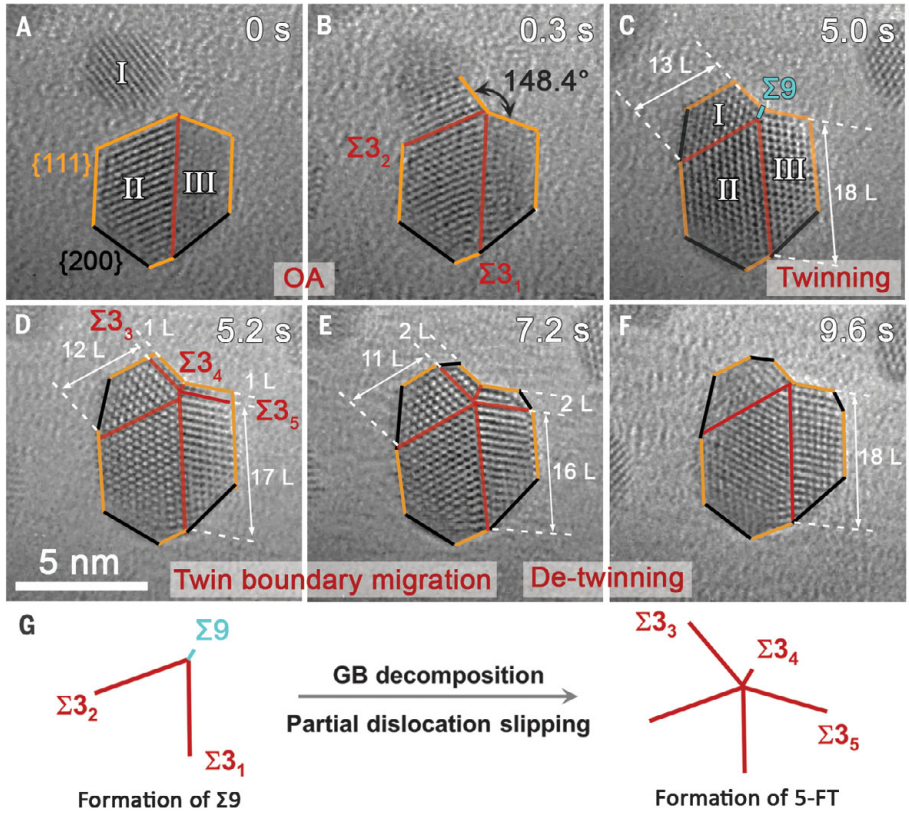


Fig. 5. Formation mechanism 2 through OA and partial dislocation slipping. (**A** and **B**) Formation of $\Sigma3_2$ and a concave surface after the OA process. (**C**) Formation of Σ9 after diffusion of atoms onto the concave surface. (**D**) Formation of 5-FT induced through GB decomposition and partial dislocation slipping. (**E** and **F**) Migration of twin boundaries. (**G**) Schematic illustration of the decomposition of Σ9 into three Σ3 GBs via partial dislocation slipping.

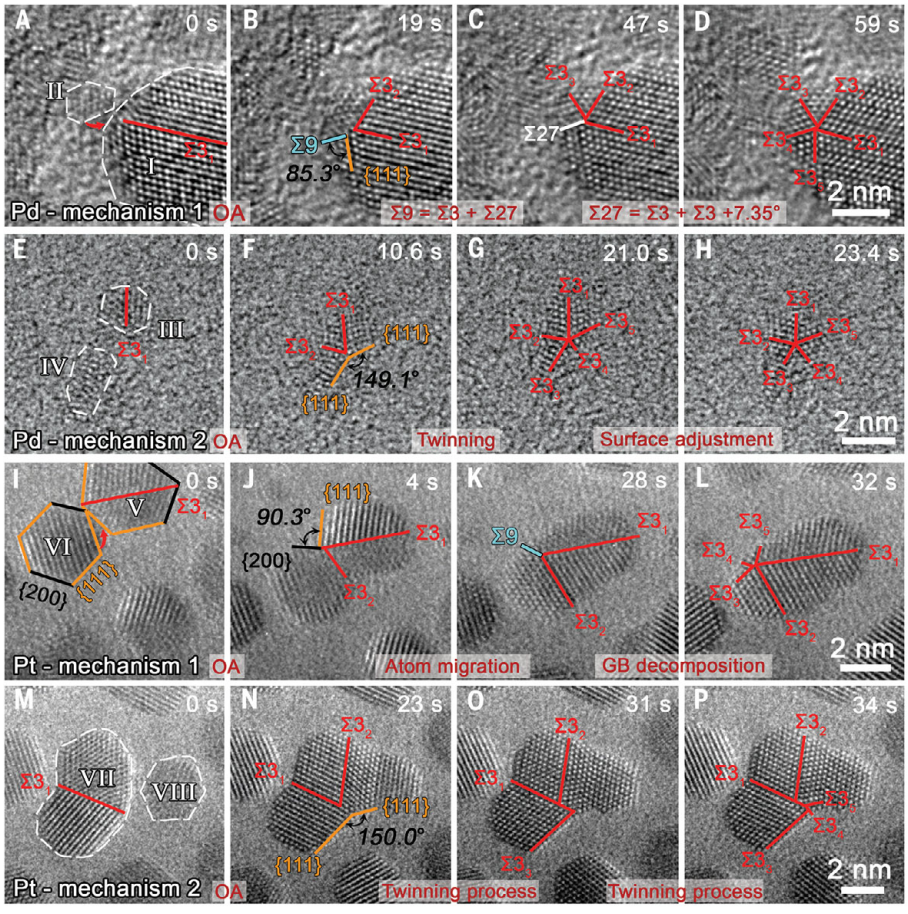


Fig. 6. Examples of formation of 5-FTs in Pd and Pt NPs by means of mechanisms 1 and 2. (**A** to **D**) Formation process of mechanism 1 in Pd NPs. (**E** to **H**) Formation process of mechanism 2 in Pd NPs. (**I** to **L**) Formation process of mechanism 1in Pt NPs. (**M** to **P**) Formation process of mechanism 2 in Pt NPs.

$\Sigma3_4$ and $\Sigma3_5$, forming 5-FT (Fig. 6, C and D) by means of mechanism 1. At a ~149° concave surface induced through OA (Fig. 6, E and F), 5-FT formed through partial dislocation slipping—that is, by means of mechanism 2 (Fig. 6, F and G).

On the basis of most observations of Pd and Pt 5-FTs, the sizes of twin units are one to three layers of {111} planes (Fig. 6 and fig. S20), which is smaller than that of Au 5-FTs (4 to 10 layers) (Fig. 2I and figs. S15 and S17). During the mechanism 1 process, we did not observe a ZST (composed of three layers of {111} planes). A ZST may or may not form depending on the size of twin units. The average particle size of the initial Pt and Pd NPs was ~2.7 nm (figs. S21 and S22), which is only slightly smaller than that of the Au NPs (~3 nm) (fig. S1). Therefore, the small twin units are unlikely to be attributable to the initial Pd or Pt particle size. For example, by means of mechanism 1, after the OA of two ~3-nm Pt NPs (Fig. 6, I and J), the twin pole of the resulting 5-FT is near the surface, with small twin units of two to three layers of {111} planes (Fig. 6, K and L). Another Pt example shows small twin units with one to two layers of {111} planes by means of mechanism 2 (Fig. 6, I to L and M to P). On the other hand, because of the higher twin-fault energy, more energy is required to form twin interfaces for Pt and Pd than Au, resulting in small twin units. For the same reason, few 5-FTs in Pt, which has the highest twin-fault energy, were reported previously (*38*). Many other factors may affect the stability of 5-FTs, such as particle morphologies, external environments, and surface absorbents (*38*). Large 5-FTs with high twin-fault energies may form under different conditions.

In total, we observed 57 (Au), 33 (Pd), and 57 (Pt) OA processes and ~26% (Au), ~15% (Pd), and ~9% (Pt) of them led to 5-FTs by means of mechanism 1 or 2, indicating that the two 5-FT formation mechanisms are applicable for various metal systems.

Discussion and outlook

Our findings reveal that OA processes lead to multiply twinned structures, among which two types of concave surfaces lead to 5-FTs through two well-defined and reproducible mechanisms. In mechanism 1, when OA leads to a ~94° concave surface with a small radius of curvature, OA provides a path to create behavior that is far from equilibrium. A large strain quickly accumulates while the surface energy is minimized through atomic surface diffusion to the curved surface, resulting in lattice deformation and high-energy grain boundaries (Fig. 3). To release this high strain energy, a small twin forms through slipping on three consecutive {111} planes, which does not create extra strain in the nearby lattice or require high energy—there is zero net strain. Continuous nucleation and growth of this twin structure releases the strain, eventually forming a 5-FT.

In mechanism 2, after surface diffusion at ~150° concave surfaces, OA only creates a small Σ9 (Fig. 5C) because of the large radius of curvature, and the crystal lattice is less deformed than in mechanism 1. Consequently, asymmetrical 5-FTs form through partial dislocation slipping of only a few layers of {111} planes because further slipping creates extra strain energy in the lattice because of large interfaces. By contrast, in mechanism 1 a highly strained lattice is already formed through surface diffusion after the OA event. The formation of 5-FTs within the bulk of the particle (with large interfaces) results in a reduction of strain. Similar processes of mechanisms 1 and 2 are also found in high twin-fault energy systems, such as Pd and Pt, but result in smaller twin units.

Multiply twinned structures, including 5-FTs, have attracted increased interest owing to their distinct properties (*6*, *7*) and have been widely used in mechanical engineering (*9*), optics (*11*), and catalysis (*12*). OA events have been recognized as a common pathway for crystal growth and occur widely in diverse systems, including oxides, metals, semiconductors, organics, and biomineral phases (*1*–*5*). Therefore, what we learned from our observation may generalize to a wide range of materials and enable the formation of multiply twinned structures besides 5-FTs, such as the formation of multiply twinned iron oxide in solution (*21*), depending on specific crystal structure and the nature of twin boundaries in each system. Our findings place disparate systems into the context of well-developed theories for multiple-twin-formation mechanisms, provide a guide for interpreting and controlling twinned crystal structures and morphologies, and hopefully will result in advances in materials design and synthesis for diverse applications.

REFERENCES AND NOTES

1. N. Daneu, H. Schmid, A. Rečnik, W. Mader, *Am. Mineral.* **92**, 1789–1799 (2007).
2. V. Šrot, A. Rečnik, C. Scheu, S. Šturm, B. Mirtič, *Am. Mineral.* **88**, 1809–1816 (2003).

3. L. Lu, X. Chen, X. Huang, K. Lu, *Science* **323**, 607–610 (2009).
4. M. R. Langille, J. Zhang, M. L. Personick, S. Li, C. A. Mirkin, *Science* **337**, 954–957 (2012).
5. C. L. Jia, K. Urban, *Science* **303**, 2001–2004 (2004).
6. J. Y. Wu, S. Nagao, J. Y. He, Z. L. Zhang, *Nano Lett.* **11**, 5264–5273 (2011).
7. Y. Li *et al*., *Nano Lett.* **17**, 1312–1317 (2017).
8. Y. Xia, Y. Xiong, B. Lim, S. E. Skrabalak, *Angew. Chem. Int. Ed.* **48**, 60–103 (2009).
9. S. Narayanan, G. Cheng, Z. Zeng, Y. Zhu, T. Zhu, *Nano Lett.* **15**, 4037–4044 (2015).
10. H. Li, H. Hu, D. Xu, *Anal. Chem.* **87**, 3826–3833 (2015).
11. B. Pietrobon, V. Kitaev, *Chem. Mater.* **20**, 5186–5190 (2008).
12. X. Wang *et al*., *J. Am. Chem. Soc.* **137**, 15036–15042 (2015).
13. G. Rose, *Ann. Phys.* **99**, 196–204 (1831).
14. H. Hofmeister, *Nanotechnology* **3**, 431–452 (2004).
15. L. D. Marks, L. Peng, *J. Phys. Condens. Matter* **28**, 053001 (2016).
16. H. Hofmeister, *Cryst. Res. Technol.* **33**, 3–25 (1998).
17. F. Baletto, A. Rapallo, G. Rossi, R. Ferrando, *Phys. Rev. B* **69**, 235421 (2004).
18. K. Yagi, K. Takayanagi, K. Kobayashi, G. Honjo, *J. Cryst. Growth* **28**, 117–124 (1975).
19. M. R. Langille, J. Zhang, C. A. Mirkin, *Angew. Chem. Int. Ed.* **50**, 3543–3547 (2011).
20. Y. T. Zhu, X. Z. Liao, R. Z. Valiev, *Appl. Phys. Lett.* **86**, 103112 (2005).
21. D. Li *et al*., *Science* **336**, 1014–1018 (2012).
22. A. S. Barnard, N. P. Young, A. I. Kirkland, M. A. van Huis, H. Xu, *ACS Nano* **3**, 1431–1436 (2009).
23. N. P. Young, M. A. van Huis, H. W. Zandbergen, H. Xu, A. I. Kirkland, *Ultramicroscopy* **110**, 506–516 (2010).
24. K. Koga, T. Ikeshoji, K. Sugawara, *Phys. Rev. Lett.* **92**, 115507 (2004).
25. Materials and methods are available as supplementary materials.
26. L. Vitos, A. V. Ruban, H. L. Skriver, J. Kollár, *Surf. Sci.* **411**, 186–202 (1998).
27. H. Grimmer, W. Bollmann, D. H. Warrington, *Acta Crystallogr. A* **30**, 197–207 (1974).
28. C. Zhu *et al*., *Nat. Commun.* **9**, 421 (2018).
29. Z. A. Peng, X. Peng, *J. Am. Chem. Soc.* **124**, 3343–3353 (2002).
30. D. Wolf, *Acta Metall. Mater.* **38**, 781–790 (1990).
31. A. Howie, L. D. Marks, *Philos. Mag. A Phys. Condens. Matter Defects Mech. Prop.* **49**, 95–109 (1984).
32. V. Randle, *Acta Mater.* **47**, 4187–4196 (1999).
33. L. Liu, J. Wang, S. K. Gong, S. X. Mao, *Phys. Rev. Lett.* **106**, 175504 (2011).
34. X. L. Ma, H. Zhou, J. Narayan, Y. T. Zhu, *Scr. Mater.* **109**, 89–93 (2015).
35. S. Patala, L. D. Marks, M. Olvera de la Cruz, *J. Phys. Chem. C* **117**, 1485–1494 (2013).
36. M. Song, Z. Wu, N. Lu, D. Li, *Chem. Mater.* **31**, 842–850 (2019).
37. N. M. Rosengaard, H. L. Skriver, *Phys. Rev. B* **47**, 12865–12873 (1993).
38. S. Zhou, M. Zhao, T. H. Yang, Y. Xia, *Mater. Today* **22**, 108–131 (2019).

ACKNOWLEDGMENTS

We thank P. Sushko and J. De Yoreo for their comments regarding revisions to our manuscript. **Funding:** This research was supported by the U.S. Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences, through Early Career Research Program Award KC0203020:67037. The work was conducted in the William R. Wiley Environmental Molecular Sciences Laboratory, which is a DOE Office of Science User Facility and located at Pacific Northwest National Laboratory (PNNL). PNNL is a multiprogram national laboratory operated for DOE by Battelle under contract DE-AC05-76RL01830. **Author contributions:** M.S. conducted experiments, data analysis, and results discussion. G.Z. conducted MD simulations. N.L. assisted with data analysis. J.L. synthesized Au NPs. E.N. assisted with image processing. H.W. assisted with MD simulations. D.L. supervised the study, designed the experiments, and carried out the results discussion. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6473/40/suppl/DC1
Materials and Methods
Figs. S1 to S25
References (*39*–*70*)
Movies S1 to S3

17 April 2019; resubmitted 7 October 2019
Accepted 15 November 2019
10.1126/science.aax6511

MULTIPLEX GENOMICS

Massively multiplex chemical transcriptomics at single-cell resolution

Sanjay R. Srivatsan[1,2]*, José L. McFaline-Figueroa[1]*, Vijay Ramani[1,3]*, Lauren Saunders[1], Junyue Cao[1], Jonathan Packer[1], Hannah A. Pliner[1], Dana L. Jackson[1], Riza M. Daza[1], Lena Christiansen[4], Fan Zhang[4], Frank Steemers[4], Jay Shendure[1,5,6,7]†‡, Cole Trapnell[1,5,7]†‡

High-throughput chemical screens typically use coarse assays such as cell survival, limiting what can be learned about mechanisms of action, off-target effects, and heterogeneous responses. Here, we introduce "sci-Plex," which uses "nuclear hashing" to quantify global transcriptional responses to thousands of independent perturbations at single-cell resolution. As a proof of concept, we applied sci-Plex to screen three cancer cell lines exposed to 188 compounds. In total, we profiled ~650,000 single-cell transcriptomes across ~5000 independent samples in one experiment. Our results reveal substantial intercellular heterogeneity in response to specific compounds, commonalities in response to families of compounds, and insight into differential properties within families. In particular, our results with histone deacetylase inhibitors support the view that chromatin acts as an important reservoir of acetate in cancer cells.

High-throughput screens (HTSs) are a cornerstone of the pharmaceutical drug-discovery pipeline (*1*, *2*). However, conventional HTSs have at least two major limitations. First, the readout of most are restricted to gross cellular phenotypes, e.g., proliferation (*3*, *4*), morphology (*5*, *6*), or a highly specific molecular readout (*7*, *8*). Subtle changes in cell state or gene expression that might otherwise provide mechanistic insights or reveal off-target effects are routinely missed.

Second, even when HTSs are performed in conjunction with more comprehensive molecular phenotyping such as transcriptional profiling (*9*–*12*), a limitation of bulk assays is that even cells ostensibly of the same "type" can exhibit heterogeneous responses (*13*, *14*). Such cellular heterogeneity can be highly relevant in vivo. For example, it remains largely unknown whether the rare subpopulations of cells that survive chemotherapeutics are doing so on the basis of their genetic background, epigenetic state, or some other aspect (*15*, *16*).

In principle, single-cell transcriptome sequencing (scRNA-seq) represents a form of high-content molecular phenotyping that could enable HTSs to overcome both limitations. However, the per-sample and per-cell costs of most scRNA-seq technologies remain high, precluding even modestly sized screens. Recently, several groups have developed "cellular hashing" methods, in which cells from different samples are molecularly labeled and mixed before scRNA-seq. However, current hashing approaches require relatively expensive reagents [e.g., antibodies (*17*) or chemically modified DNA oligos (*18*, *19*)], use cell-type–dependent protocols (*20*), and/or use scRNA-seq platforms with a high per-cell cost.

To enable cost-effective HTSs with scRNA-seq–based phenotyping, we describe a new sample labeling (hashing) strategy that relies on labeling nuclei with unmodified single-stranded DNA oligos. Recent improvements in single-cell combinatorial indexing (sci-RNA-seq3) have lowered the cost of scRNA-seq library preparation to <$0.01 per cell, with millions of cells profiled per experiment (*21*). Here, we combine nuclear hashing and sci-RNA-seq into a single workflow for multiplex transcriptomics in a process called "sci-Plex." As a proof of concept, we use sci-Plex to perform HTS on three cancer cell lines, profiling thousands of independent perturbations in a single experiment. We further explore how chemical transcriptomics at single-cell resolution can shed light on mechanisms of action. Most notably, we find that gene-regulatory changes consequent to treatment with histone deacetylase (HDAC) inhibitors are consistent with the model that they interfere with proliferation by restricting a cell's ability to draw acetate from chromatin (*22*, *23*).

Results

Nuclear hashing enables multisample sci-RNA-seq

Single-cell combinatorial indexing (sci-) methods use split-pool barcoding to specifically label the molecular contents of large numbers of single cells or nuclei (*24*). Samples can be barcoded by these same indices, e.g., by placing each sample in its own well during reverse transcription in sci-RNA-seq (*21*, *25*), but such enzymatic labeling at the scale of thousands of samples is operationally infeasible and cost prohibitive. To enable single-cell molecular profiling of a large number of independent samples within a single sci- experiment, we set out to develop a low-cost labeling procedure.

We noticed that single-stranded DNA (ssDNA) specifically stained the nuclei of permeabilized cells but not intact cells (Fig. 1A and fig. S1A). We therefore postulated that a polyadenylated ssDNA oligonucleotide could be used to label populations of nuclei in a manner compatible with sci-RNA-seq (Fig. 1B and fig. S1B). To test this concept, we performed a "barnyard" experiment. We separately seeded human (HEK293T) and mouse (NIH3T3) cells to 48 wells of a 96-well culture plate. We then performed nuclear lysis in the presence of 96 well-specific polyadenylated ssDNA oligos ("hash oligos") and fixed the resulting nuclear suspensions with paraformaldehyde. Having labeled or "hashed" the nuclei with a molecular barcode, we pooled nuclei and performed a two-level sci-RNA-seq experiment. Because the hash oligos were polyadenylated, they had the potential to be combinatorially indexed identically to endogenous mRNAs. As intended, we recovered reads corresponding to both endogenous mRNAs [median 4740 unique molecular identifiers (UMIs) per cell] and hash oligos (median 270 UMIs per cell).

We devised a statistical framework to identify the hash oligos associated with each cell at a frequency exceeding background (table S1). We observed 99.1% concordance between species assignments on the basis of hash oligos versus endogenous cellular transcriptomes (Fig. 1C and fig. S1, C to F). Additionally, the association of hash oligos and nuclei was stable to a freeze–thaw cycle, highlighting the opportunity to label and store samples (Fig. 1D and fig. S1, G and H). These results demonstrate that hash oligos stably label nuclei in a manner that is compatible with sci-RNA-seq.

In sci- experiments, "collisions" are instances in which two or more cells are labeled with the same combination of barcodes by chance (*24*). To evaluate hashing as a means of detecting doublets resulting from collisions, we varied the number of nuclei loaded per polymerase chain reaction well, resulting in a range of predicted collision rates (7 to 23%) that was well matched by observation (fig. S1I). Hash oligos facilitated the identification of the vast majority of interspecies doublets (95.5%) and otherwise undetectable within-species doublets (Fig. 1E and fig. S1, J and K).

sci-Plex enables multiplex chemical transcriptomics at single-cell resolution

We next evaluated whether nuclear hashing could enable chemical screens by labeling cells

[1]Department of Genome Sciences, University of Washington, Seattle, WA, USA. [2]Medical Scientist Training Program, University of Washington, Seattle, WA, USA. [3]Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA, USA. [4]Illumina Inc., San Diego, CA, USA. [5]Allen Discovery Center for Cell Lineage Tracing, Seattle, WA, USA. [6]Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA. [7]Brotman Baty Institute for Precision Medicine, Seattle, WA, USA.
*These authors contributed equally to this work. †These authors contributed equally to this work.
‡Corresponding author. Email: shendure@uw.edu (J.S.); coletrap@uw.edu (C.T.)

that had undergone a specific perturbation, followed by single-cell transcriptional profiling as a high-content phenotypic assay. We exposed A549, a human lung adenocarcinoma cell line, to one of four compounds: dexamethasone (a corticosteroid agonist), nutlin-3a (a p53-Mdm2 antagonist), BMS-345541 (an inhibitor of nuclear factor κB–dependent transcription), or vorinostat [suberoylanilide hydroxamic acid (SAHA), an HDAC inhibitor], for 24 hours across seven doses in triplicate for a total of 84 drug–dose–replicate combinations and additional vehicle controls (Fig. 2A and fig. S2A). We labeled nuclei from each well and subjected them to sci-RNA-seq2 (fig. S2, B to D, and table S1).

We used Monocle 3 (*21*) to visualize these data using Uniform Manifold Approximation and Projection (*26*) (UMAP) and Louvain community detection to identify compound-specific clusters of cells, which were distributed in a dose-dependent manner (Fig. 2, B and C, and fig. S2, E and F). To quantify the "population average" transcriptional response of A549 cells to each of the four drugs, we modeled each gene's expression as a function of dose through generalized linear regression. A total of 7561 genes were sensitive to at least one drug, and 3189 genes were differentially expressed in response to multiple drugs (fig. S3A and table S2). These included canonical targets of dexamethasone (Fig. 2D) and nutlin-3a (Fig. 2E). Gene ontology analysis of differentially expressed genes revealed the involvement of drug-specific pathways (e.g., hormone signaling for dexamethasone; p53 signaling for nutlin-3a; fig. S3B). Additionally, we evaluated whether the number of cells recovered at each concentration could be used to infer toxicity akin to traditional screens. After fitting a response curve to the recovered cellular counts, we inferred a "viability score" from sci-Plex data, a metric that was concordant with "gold standard" measurements (Fig. 2F and fig. S2, G to I).

sci-Plex scales to thousands of samples and enables HTS

To assess how sci-Plex scales for HTS, we performed a screen of 188 compounds targeting a diverse range of enzymes and molecular pathways (Fig. 3A). Half of this panel was chosen to target transcriptional and epigenetic regulators. The other half was chosen to sample diverse mechanisms of action. We exposed three well-characterized human cancer cell lines, A549 (lung adenocarcinoma), K562 (chronic myelogenous leukemia), and MCF7 (mammary adenocarcinoma), to each of these 188 compounds at four doses (10 nM, 100 nM, 1 μM, and 10 μM) in duplicate, randomizing compounds and doses across well positions in replicate culture plates (table S3). These conditions, together with vehicle controls, accounted for 4608 of 4992 independently treated cell populations in this experiment. After treatment, we lysed cells to expose nuclei, hashed them with a specific combination of two oligos (fig. S4A), and performed sci-RNA-seq3 (*21*). After sequencing and filtering based on hash purity (fig. S4, B to F), we obtained transcriptomes for 649,340 single cells, with median mRNA UMI counts of 1271, 1071, and 2407 for A549, K562, and MCF7, respectively (fig. S5A). The aggregate expression profiles for each cell type were highly concordant between replicate wells (Pearson correlation = 0.99) (fig. S5B).

Visualizing sci-RNA-seq profiles separately for each cell line revealed compound-specific transcriptional responses and patterns that were common to multiple compounds. For each of the cell lines, UMAP projected most cells into a central mass, flanked by smaller clusters (Fig. 3B). These smaller clusters were largely composed of cells treated with compounds from only one or two compound classes (figs. S6 and S7, A to C). For example, A549 cells treated with triamcinolone acetonide, a synthetic glucocorticoid receptor agonist, were markedly enriched in one such small cluster, comprising 95% of its cells [Fisher's exact test, false discovery rate (FDR) < 1%; fig. S7, D and E]. Although many drugs were associated with a seemingly homogeneous transcriptional response, we also identified cases in which distinct transcriptional states were induced by the same drug. For example, in A549, the microtubule-stabilizing compounds epothilone A and epothilone B were associated with three such focal enrichments, each composed of cells from both compounds at all four doses (fig. S7, F and G). The cells in each focus were distinct from one another, but transcriptionally similar to other treatments: a recently identified

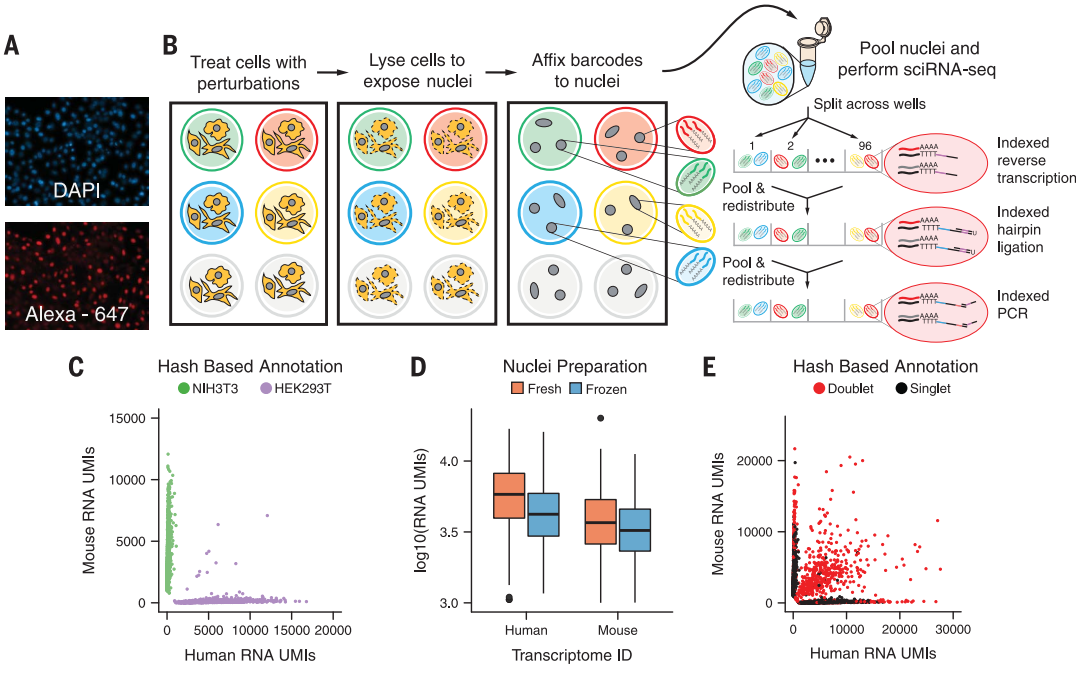


Fig. 1. sci-Plex uses polyadenylated single-stranded oligonucleotides to label nuclei, enabling cell hashing and doublet detection. (**A**) Fluorescent images of permeabilized nuclei after incubation with DAPI (top) and an Alexa Fluor-647–conjugated single-stranded oligonucleotide (bottom). (**B**) Overview of sci-Plex. Cells corresponding to different perturbations are lysed in-well, their nuclei labeled with well-specific "hash" oligos, followed by fixation, pooling, and sci-RNA-seq. (**C**) Scatter plot depicting the number of UMIs from single-cell transcriptomes derived from a mixture of hashed human HEK293T cells and murine NIH3T3 cells. Points are colored on the basis of hash oligo assignment. (**D**) Boxplot depicting the number of mRNA UMIs recovered per cell for fresh versus frozen human and mouse cell lines. (**E**) Scatter plot of overloading experiment; axes are as in (C). Identified hash oligo collisions (red) identify cellular collisions with high sensitivity.

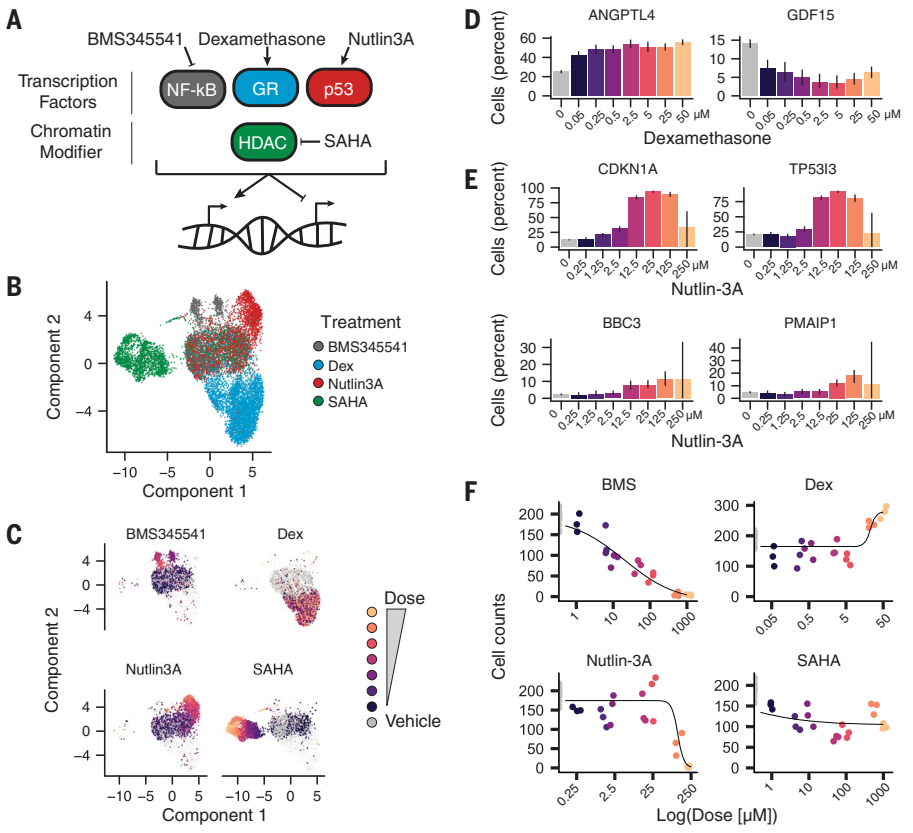


Fig. 2. sci-Plex enables multiplex chemical transcriptomics at single-cell resolution. (**A**) Diagram depicting compounds and corresponding targets assayed within the pilot sci-Plex experiment. A549 lung adenocarcinoma cells were treated with either vehicle [dimethylsulfoxide (DMSO) or ethanol] or one of four compounds (BMS345541, dexamethasone, nutlin-3a, or SAHA). (**B**) UMAP embedding of chemically perturbed A549 cells colored by drug treatment. (**C**) UMAP embedding of chemically perturbed A549 cells faceted by treatment with cells colored by dose. (**D** and **E**) Expression of a canonical (D) glucocorticoid receptor activated (*ANGPTL4*) and repressed (*GDF15*) target genes as a function of dexamethasone dose or (E) p53 target genes as a function of nutlin-3a dose. *y*-axes indicate the percentage of cells with at least one read corresponding to the transcript. (**F**) Dose–response viability estimates for BMS345541-, dexamethasone-, nutlin-3a-, and SAHA-treated A549 cells on the basis of the relative number of cells recovered at each dose.

microtubule destabilizer, rigosertib (*27*); the SETD8 inhibitor UNC0397; or untreated proliferating cells (fig. S7H).

We next assessed the effects of each drug on the "population average" transcriptome of each cell line. In total, 6238 genes were differentially expressed in a dose-dependent manner in at least one cell line (FDR < 5%; fig. S8 and tables S4 and S5). Bulk RNA-seq measurements collected for five compounds across four doses and vehicle agreed with averaged gene expression values and estimated effect sizes across identically treated single cells, although correlations between small effect sizes were diminished (fig. S9). Moreover, sci-Plex dose-dependent effect profiles correlated with compound-matched L1000 measurements (*11*) (fig. S10).

Genes associated with the cell cycle were highly variable across individual cells, and many drugs reduced the fraction of cells that expressed proliferation marker genes (figs. S11 and S12). In principle, scRNA-seq should be able to distinguish shifts in the proportion of cells in distinct transcriptional states from gene-regulatory changes within those states. By contrast, bulk transcriptome profiling would confound these two signals (fig. S13A) (*14*). We therefore tested for dose-dependent differential expression on subsets of cells corresponding to the same drug but expressing high versus low levels of proliferation marker genes (fig. S13B). Correlation between the dose-dependent effects on the two fractions of each cell type varied across drug classes (fig. S13C), with some frankly discordant effects for individual compounds (fig. S13D). Viability analysis performed as in the pilot experiment revealed that after drug exposure at the highest dose, only 52 (27%) compounds caused a decrease in viability of 50% or more (Fig. 3C and fig. S5C). Among the drugs that reduced viability, we observed a higher sensitivity of K562 to the Src and Abl inhibitor bosutinib (Fig. 3C), a result that we confirmed by cell counting (fig. S14A). This result is consistent with K562 cells harboring a constitutively active BCR-ABL fusion kinase (*28*) and an observed increased sensitivity of hematopoietic and lymphoid cancer cell lines to Abl inhibitors (*29*) (fig. S14B).

To assess whether each compound elicited similar responses across the three cell lines, we clustered compounds using the effect sizes for dose-dependent genes as loadings in each cell line (figs. S15 to S18). Joint analysis of the three cell lines revealed common and cell-type–specific responses to different compounds (figs. S19 and S20). For example, trametinib, a mitogen-activated protein kinase kinase (MEK) inhibitor, induced a transcriptionally distinct response in MCF7 cells. Inspection of UMAP projections revealed trametinib-treated MCF7 cells interspersed among vehicle controls, reflecting limited effects. By contrast, trametinib-treated A549 and K562 cells, which harbor activating KRAS and ABL mutations (*30*), respectively, were tightly clustered, consistent with a strong, specific transcriptional response to inhibition of MEK signaling by trametenib (Fig. 3D). Further, we observed that these A549 and K562 cells appeared proximal to clusters enriched with inhibitors of HSP90, a key chaperone for protein folding (Fig. 3D). This observation was corroborated by concordant changes in HSP90AA1 expression in trametinib-treated cells (Fig. 3E). Analysis of Connectivity Map data (*11*, *12*) revealed further evidence that MEK inhibitors do indeed induce highly similar gene expression signatures to HSP90 perturbations (fig. S14C), especially in A549 but not in MCF7 (fig. S14, D and E). These results are concordant with previous observations of the regulation of HSP90AA1 downstream of MEK signaling (*31*) and suggest that similarity in single-cell transcriptomes treated with distinct compounds can highlight drugs that target convergent molecular pathways.

Inference of chemical and mechanistic properties of HDAC inhibitors

For each of the three cell lines, the most prominent compound response was composed of cells treated with one of 17 HDAC inhibitors (Fig. 3B, dark blue, and table S6). To assess the similarity of the dose–response trajectories between cell lines, we aligned HDAC-treated cells and vehicle-treated cells from all three cell lines using a mutual-nearest neighbor (MNN) matching approach (*32*) to produce a consensus HDAC inhibitor trajectory, which we call "pseudodose" [analogous to "pseudotime" (*33*)] (Fig. 4A and fig. S21). We observed that some HDAC inhibitors induced homogeneous responses, with nearly all cells localized to a relatively narrow range of the HDAC inhibitor trajectory at each dose (e.g., pracinostat in

A549), whereas other drugs induced much greater cellular heterogeneity (Fig. 4B and fig. S22).

Such heterogeneity could be explained by cells executing a defined transcriptional program asynchronously, with the dose of drug that the cells are exposed to modulating the rates of their progression through it. To test this hypothesis, we sequenced the transcriptomes of 64,440 A549 cells that were treated for 72 hours with one of 48 compounds, including many of the HDAC inhibitors from the large sci-Plex screen. Upon accounting for confluency-dependent cell-cycle effects and MNN alignment (figs. S23 and S24), the co-embedded UMAP projection revealed new focal concentrations of cells at 72 hours that were not evident at the 24-hour time point, e.g., SRT1024 (fig. S25). However, for the majority of HDAC inhibitors tested, we did not observe that cells at a given dose moved

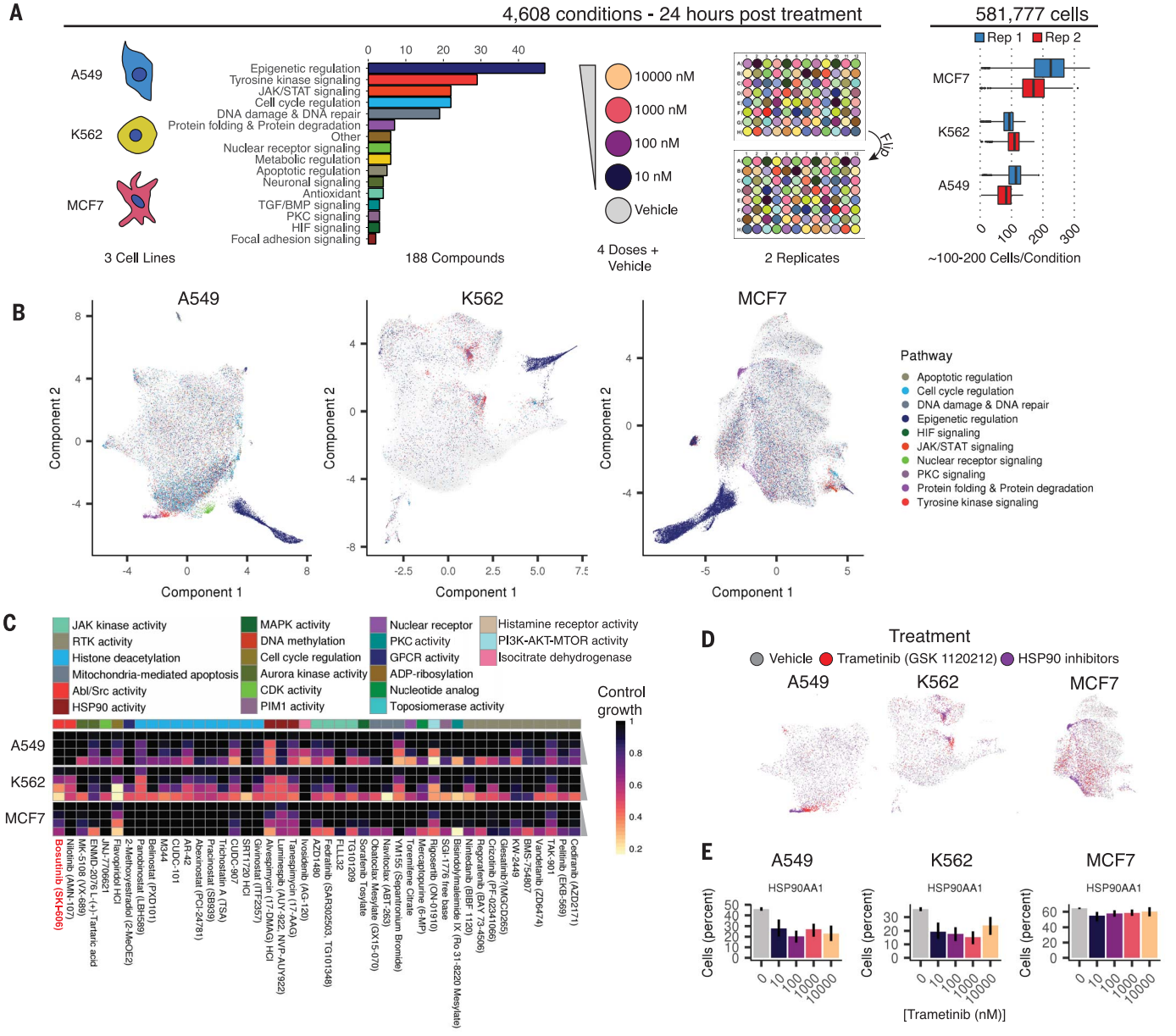


Fig. 3. sci-Plex enables global transcriptional profiling of thousands of chemical perturbations in a single experiment. (**A**) Schematic of the large-scale sci-Plex experiment (sci-RNA-seq3). A total of 188 small molecules were tested for their effects on A549, K562, and MCF7 human cell lines, each at four doses and in biological replicate, after 24 hours of treatment. The plate positions of doses and drugs were varied between replicates, and a median of 100 to 200 cells were recovered per condition. Colors demarcate cell line, compound pathway, and dose. (**B**) UMAP embeddings of A549, K562, and MCF7 cells in our screen with each cell colored by the pathway targeted by the compound to which a given cell was exposed. To facilitate visualization of significant molecular phenotypes, we added transparency to cells treated with compound or dose combinations that did not appreciably alter the corresponding cells' distribution in UMAP space compared with vehicle controls (Fisher's exact test, FDR < 1%). (**C**) Viability estimates obtained from hash-based counts of nuclei at each dose of selected compounds (bosutinib is highlighted in red text). Rows represent compound doses increasing from top to bottom, and columns represent individual compounds. Annotation bar at top depicts the broad cellular activity targeted by each compound. (**D**) UMAP embeddings highlighted by treatment with the MEK inhibitor trametinib (red), an HSP90 inhibitor (purple), or vehicle control (gray). (**E**) *HSP90AA1* expression levels in cells exposed to increasing doses of trametinib. *y*-axes indicate the percentage of cells with at least one read corresponding to the transcript.

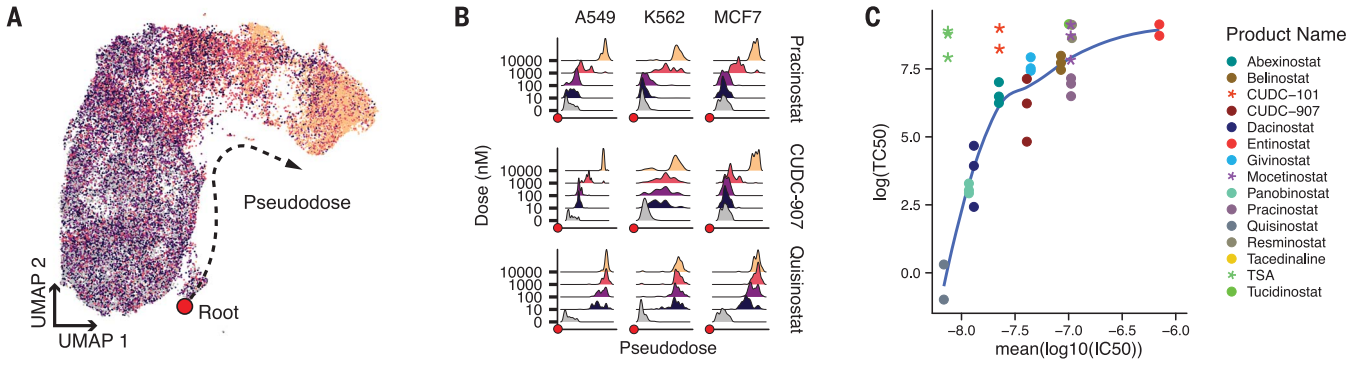


Fig. 4. HDAC inhibitor trajectory captures cellular heterogeneity in drug response and biochemical affinity. (**A**) MNN alignment and UMAP embedding of transcriptional profiles of cells treated with one of 17 HDAC inhibitors. Pseudodose root is displayed as a red dot. (**B**) Ridge plots displaying the distribution of cells along pseudodose by dose shown for three HDAC inhibitors with varying biochemical affinities. (**C**) Relationship between TC_{50} and average $\log_{10}(IC_{50})$ from in vitro measurements. Asterisks indicate compounds with a solubility <200 mM (in DMSO) that were not included in the fit.

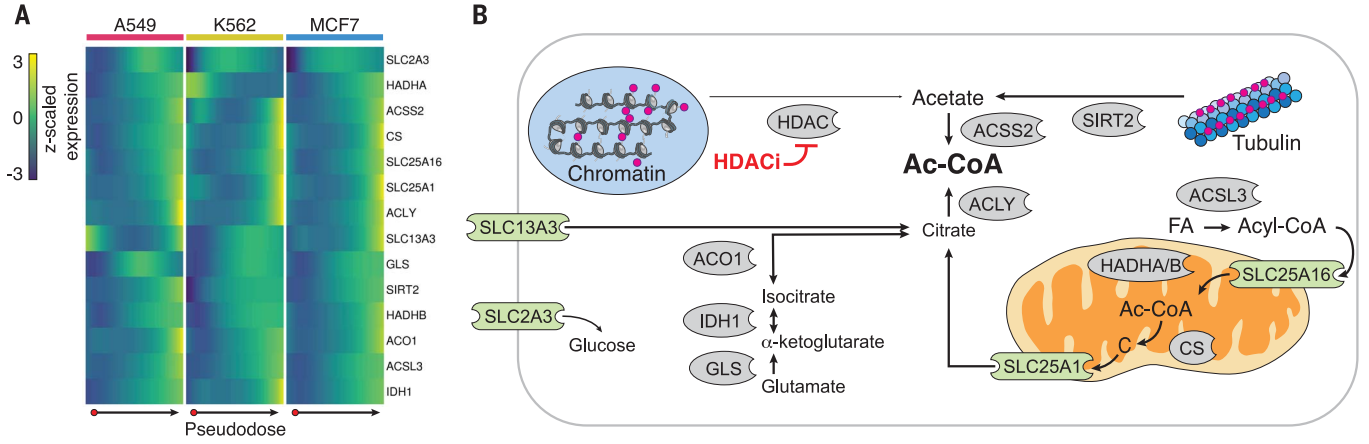


Fig. 5. HDAC inhibitors shared transcriptional response indicative of acetyl-CoA deprivation. (**A**) Heatmap of row-centered and *z*-scaled gene expression depicting the up-regulation of pseudodose-dependent genes involved in cellular carbon metabolism. (**B**) Diagram of the roles of genes from (A) in cytoplasmic acetyl-CoA regulation. Red circles indicate acetyl groups. Enzymes are shown in gray. Transporters are shown in green (FA, fatty acid; Ac-CoA, acetyl-CoA; C, citrate).

farther along an aligned HDAC trajectory at 72 hours (fig. S26). This suggests that the dose of many HDAC inhibitors governs the magnitude of a cell's response rather than its rate of progression and that any observed heterogeneity cannot be attributed solely to asynchrony (fig. S26).

Next, we assessed whether a given HDAC inhibitor's target affinity explained its global transcriptional response to the compound. We used dose-response models to estimate each compound's transcriptional median effective concentration (TC_{50}), i.e., the concentration needed to drive a cell halfway across the HDAC inhibitor pseudodose trajectory (fig. S27A and table S6). To compare the transcriptionally derived measures of potency with the biochemical properties of each compound, we collected published median inhibitory concentration (IC_{50}) values for each compound from in vitro assays performed on eight purified HDAC isoforms (table S7). With the exception of two relatively insoluble compounds, our calculated TC_{50} values increased as a function of compound IC_{50} values (Fig. 4C and fig. S27, B and C).

To assess the components of the HDAC inhibitor trajectory, we performed differential expression analysis using pseudodose as a continuous covariate. Of the 4308 genes that were significantly differentially expressed over this consensus trajectory, 2081 (48%) responded in a cell-type–dependent manner and 942 (22%) exhibited the same pattern in all three cell lines (fig. S28, A and B, and table S8). One prominent pattern shared by the three cell lines was an enrichment for genes and pathways indicative of progression toward cell-cycle arrest (figs. S28C and S29, A and B). DNA content staining and flow cytometry confirmed that HDAC inhibition resulted in the accumulation of cells in the G_2/M phase of the cell cycle (*34*) (fig. S29, C and D).

The shared response to HDAC inhibition included not only cell-cycle arrest but also the altered expression of genes involved in cellular metabolism (fig. S28C). Histone acetyltransferases and deacetylases regulate chromatin accessibility and transcription factor activity through the addition or removal of charged acetyl groups (*35*–*37*). Acetate, the product of HDAC class I-, II-, and IV-mediated histone deacetylation and a precursor to acetyl-coenzyme A (acetyl-CoA), is required for histone acetylation but also has important roles in metabolic homeostasis (*23*, *38*, *39*). Inhibition of nuclear deacetylation limits recycling of chromatin-bound acetyl groups for both catabolic and anabolic processes (*39*). Accordingly, we observed that HDAC inhibition led to sequestration of acetate in the form of markedly increased acetylated lysine levels after exposure to a 10 μM dose of the HDAC inhibitors pracinostat and abexinostat (fig. S30).

Upon further inspection of pseudodose-dependent genes, we observed that enzymes critical for cytoplasmic acetyl-CoA synthesis from either citrate (ACLY) or acetate (ACSS2) were up-regulated (Fig. 5A). Genes involved in

cytoplasmic citrate homeostasis (GLS, IDH1, and ACO1), citrate cellular import (SLC13A3), and mitochondrial citrate production and export (CS, SLC25A1) were also up-regulated. Up-regulation of SIRT2, which deacetylates tubulin, was also observed in response to HDAC inhibition.

Together with increases in chromatin-bound acetate, these transcriptional responses suggest a metabolically consequential depletion of cellular acetyl-CoA reserves in HDAC-inhibited cells (Fig. 5B). To validate this further, we sought to shift the distribution of cells along the HDAC inhibitor trajectory by modulating cellular acetyl-CoA levels. We treated A549 and MCF7 cells with pracinostat in the presence and absence of acetyl-CoA precursors (acetate, pyruvate, or citrate) or inhibitors of enzymes (ACLY, ACSS2, or PDH) involved in replenishing acetyl-CoA pools. After treatment, cells were harvested and processed using sci-Plex and trajectories constructed for each cell line (figs. S31 and S32). In both A549 and MCF7 cells, acetate, pyruvate, and citrate supplementation was capable of blocking pracinostat-treated cells from reaching the end of the HDAC inhibitor trajectory (fig. S31, F, J, H, and L). In MCF7 cells, both ACLY and ACSS2 inhibition shifted cells farther along the HDAC inhibitor trajectory, although no such shift was observed in A549 (fig. S31, G, K, I, and M). Taken together, these results suggest that a major feature of the response of cells to HDAC inhibitors, and possibly their associated toxicity, is the induction of an acetyl-CoA–deprived state.

Discussion

Here, we present sci-Plex, a massively multiplex platform for single-cell transcriptomics. sci-Plex uses chemical fixation to cost-effectively and irreversibly label nuclei with short, unmodified ssDNA oligos. In the proof-of-concept experiment described here, we applied sci-Plex to quantify the dose-dependent responses of cancer cells to 188 compounds through an assay that is both high content (global transcription) and high resolution (single cell). By profiling several distinct cancer cell lines, we distinguished between shared and cell-line–specific molecular responses to each compound.

sci-Plex offers some distinctive advantages over conventional HTS: it can distinguish a compound's distinct effects on cellular subsets (including complex in vitro systems such as cellular reprogramming, organoids, and synthetic embryos); it can unmask heterogeneity in cellular response to a perturbation; and it can measure how drugs shift the relative proportions of transcriptionally distinct subsets of cells. Highlighting these features, our study provides insight into the mechanism of action of HDAC inhibitors. Specifically, we find that the main transcriptional responses to HDAC inhibitors involve cell-cycle arrest and marked shifts in genes related to acetyl-CoA metabolism. For some HDAC inhibitors, we observed clear heterogeneity in responses observed at the single-cell level. Although HDAC inhibition is conventionally thought to act through mechanisms directly involving chromatin regulation, our data support an alternative model, albeit not a mutually exclusive one, in which HDAC inhibitors impair growth and proliferation by interfering with a cancer cell's ability to draw acetate from chromatin (*22*, *23*, *39*). As such, variation in cells' acetate reservoirs is a potential explanation for their heterogeneous responses to HDAC inhibitors.

As the cost of single-cell sequencing continues to fall, the opportunities for leveraging sci-Plex for basic and applied goals in biomedicine may be substantial. The proof-of-concept experiments described here, consisting of nearly 5000 independent treatments with transcriptional profiling of >100 single cells per treatment, can potentially be scaled toward a comprehensive, high-resolution atlas of cellular responses to pharmacologic perturbations (e.g., hundreds of cell lines or genetic backgrounds, thousands of compounds, multichannel single-cell profiling, etc.). The ease and low cost of oligo hashing, coupled with the flexibility and exponential scalability of single-cell combinatorial indexing, would facilitate this goal.

REFERENCES AND NOTES

1. J. R. Broach, J. Thorner, *Nature* **384** (Suppl), 14–16 (1996).
2. D. A. Pereira, J. A. Williams, *Br. J. Pharmacol.* **152**, 53–61 (2007).
3. D. Shum *et al.*, *J. Enzyme Inhib. Med. Chem.* **23**, 931–945 (2008).
4. C. Yu *et al.*, *Nat. Biotechnol.* **34**, 419–423 (2016).
5. Z. E. Perlman *et al.*, *Science* **306**, 1194–1198 (2004).
6. Y. Futamura *et al.*, *Chem. Biol.* **19**, 1620–1630 (2012).
7. J. Kang *et al.*, *Nat. Biotechnol.* **34**, 70–77 (2016).
8. K. L. Huss, P. E. Blonigen, R. M. Campbell, *J. Biomol. Screen.* **12**, 578–584 (2007).
9. C. Ye *et al.*, *Nat. Commun.* **9**, 4307 (2018).
10. E. C. Bush *et al.*, *Nat. Commun.* **8**, 105 (2017).
11. A. Subramanian *et al.*, *Cell* **171**, 1437–1452.e17 (2017).
12. J. Lamb *et al.*, *Science* **313**, 1929–1935 (2006).
13. M. B. Elowitz, A. J. Levine, E. D. Siggia, P. S. Swain, *Science* **297**, 1183–1186 (2002).
14. C. Trapnell, *Genome Res.* **25**, 1491–1498 (2015).
15. S. M. Shaffer *et al.*, *Nature* **546**, 431–435 (2017).
16. S. L. Spencer, S. Gaudet, J. G. Albeck, J. M. Burke, P. K. Sorger, *Nature* **459**, 428–432 (2009).
17. M. Stoeckius *et al.*, *Genome Biol.* **19**, 224 (2018).
18. J. Gehring, J. H. Park, S. Chen, M. Thomson, L. Pachter, Highly multiplexed single-cell RNA-seq for defining cell population and transcriptional spaces. bioRxiv 315333 [Preprint] 5 May 2018. doi.org/10.1101/315333.
19. C. S. McGinnis *et al.*, *Nat. Methods* **16**, 619–626 (2019).
20. D. Shin, W. Lee, J. H. Lee, D. Bang, *Sci. Adv.* **5**, eaav2249 (2019).
21. J. Cao *et al.*, *Nature* **566**, 496–502 (2019).
22. M. A. McBrian *et al.*, *Mol. Cell* **49**, 310–321 (2013).
23. S. A. Comerford *et al.*, *Cell* **159**, 1591–1602 (2014).
24. D. A. Cusanovich *et al.*, *Science* **348**, 910–914 (2015).
25. J. Cao *et al.*, *Science* **357**, 661–667 (2017).
26. L. McInnes, J. Healy, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv:1802.03426 [stat.ML] (9 February 2018).
27. M. Jost *et al.*, *Mol. Cell* **68**, 210–223.e6 (2017).
28. G. Grosveld *et al.*, *Mol. Cell. Biol.* **6**, 607–616 (1986).
29. E. K. Greuber, P. Smith-Pearson, J. Wang, A. M. Pendergast, *Nat. Rev. Cancer* **13**, 559–571 (2013).
30. J. Barretina *et al.*, *Nature* **483**, 603–607 (2012).
31. C. Dai *et al.*, *J. Clin. Invest.* **122**, 3742–3754 (2012).
32. L. Haghverdi, A. T. L. Lun, M. D. Morgan, J. C. Marioni, *Nat. Biotechnol.* **36**, 421–427 (2018).
33. C. Trapnell *et al.*, *Nat. Biotechnol.* **32**, 381–386 (2014).
34. W. Brazelle *et al.*, *PLOS ONE* **5**, e14335 (2010).
35. J.-S. Roe, F. Mercan, K. Rivera, D. J. Pappin, C. R. Vakoc, *Mol. Cell* **58**, 1028–1039 (2015).
36. J. E. Brownell *et al.*, *Cell* **84**, 843–851 (1996).
37. J. Taunton, C. A. Hassig, S. L. Schreiber, *Science* **272**, 408–411 (1996).
38. S. K. Kurdistani, *Curr. Opin. Genet. Dev.* **26**, 53–58 (2014).
39. K. E. Wellen *et al.*, *Science* **324**, 1076–1080 (2009).

ACKNOWLEDGMENTS

We thank members of the Shendure lab; the Trapnell lab; and others, particularly A. Adey, G. Booth, A. Hill, S. Henikoff, K. Cherukumilli, P. Selvaraj, and T. Zhou, for helpful suggestions, discussion, and mentorship; and D. Prunkard and A. Leith for assistance in flow sorting. **Funding:** This work was funded by grants from the NIH (DP1HG007811 and R01HG006283 to J.S.; DP2 HD088158 to C.T.), the W. M. Keck Foundation (to C.T. and J.S.), the NSF (DGE-1258485 to S.R.S.), and the Paul G. Allen Frontiers Group (to J.S. and C.T.). J.S. is an Investigator of the Howard Hughes Medical Institute. **Author contributions:** S.R.S., J.L.M.-F., V.R., J.S., and C.T. conceived the project; S.R.S., J.L.M.-F., and V.R. designed experiments; F.Z., L.C., and F.S. provided reagents and performed sequencing; S.R.S., J.L.M.-F., and C.T. analyzed the data; and S.R.S., J.L.M.-F., V.R., J.S., and C.T. wrote the manuscript. **Competing interests:** L.C., F.Z., and F.S. declare competing financial interests in the form of stock ownership and paid employment by Illumina, Inc. One or more embodiments of one or more patents and patent applications filed by Illumina and the University of Washington may encompass the methods, reagents, and data disclosed in this manuscript. **Data and materials availability:** Processed and raw data can be downloaded from NCBI GEO (#GSE139944). Code used to perform analyses can be accessed on Zenodo and https://github.com/cole-trapnell-lab/sci-plex. All methods for making the transposase complexes used in this paper are described in Cao *et al.* (*25*); however, Illumina will provide transposase complexes in response to reasonable requests from the scientific community subject to a material transfer agreement.

SUPPLEMENTARY MATERIAL

science.sciencemag.org/content/367/6473/45/suppl/DC1
Materials and Methods
Figs. S1 to S32
Tables S1 to S8
References (*40*–*45*)

View/request a protocol for this paper from *Bio-protocol*.

8 April 2019; accepted 18 November 2019
Published online 5 December 2019
10.1126/science.aax6234

MALARIA

A Kelch13-defined endocytosis pathway mediates artemisinin resistance in malaria parasites

Jakob Birnbaum[1]*†, Sarah Scharf[1]†, Sabine Schmidt[1], Ernst Jonscher[1], Wieteke Anna Maria Hoeijmakers[2], Sven Flemming[1]‡, Christa Geeke Toenhake[2], Marius Schmitt[1], Ricarda Sabitzki[1], Bärbel Bergmann[1], Ulrike Fröhlke[1], Paolo Mesén-Ramírez[1], Alexandra Blancke Soares[1], Hendrik Herrmann[1]§, Richárd Bártfai[2], Tobias Spielmann[1]¶

Artemisinin and its derivatives (ARTs) are the frontline drugs against malaria, but resistance is jeopardizing their effectiveness. ART resistance is mediated by mutations in the parasite's Kelch13 protein, but Kelch13 function and its role in resistance remain unclear. In this study, we identified proteins located at a Kelch13-defined compartment. Inactivation of eight of these proteins, including Kelch13, rendered parasites resistant to ART, revealing a pathway critical for resistance. Functional analysis showed that these proteins are required for endocytosis of hemoglobin from the host cell. Parasites with inactivated Kelch13 or a resistance-conferring Kelch13 mutation displayed reduced hemoglobin endocytosis. ARTs are activated by degradation products of hemoglobin. Hence, reduced activity of Kelch13 and its interactors diminishes hemoglobin endocytosis and thereby ART activation, resulting in parasite resistance.

Life-threatening malaria is caused by *Plasmodium falciparum* parasites continuously multiplying in human red blood cells. Artemisinin and its derivatives (ARTs) are the recommended first-line drugs against malaria (*1*), but their usefulness is threatened by "resistance" (*2*), defined as reduced clearance of parasites in ART-treated malaria patients (*3*). ART resistance manifests as a decreased susceptibility of young ring-stage parasites to a short pulse of ART, a property linked to delayed clearance of parasites and recrudescence of disease in patients (*4*–*6*). This can be measured in vitro using a ring-stage survival assay (RSA) (*4*).

ART resistance is primarily associated with point mutations in the parasite's Kelch propeller protein Kelch13 (*7*, *8*). At present, there is no coherent picture of the mechanism of ART resistance in the parasite, which, among others, has been connected to increased cellular stress, an activated unfolded protein response, reduced protein translation, altered DNA replication, and increased levels of phosphatidylinositol 3-phosphate (*9*–*16*). Conditional inactivation of Kelch13 causes an arrest of parasites in ring stages (*17*). However, the cellular function of Kelch13 and its involvement in ART resistance remain enigmatic. Here, we study Kelch13 and identify its interactors and their function. This reveals an entire pathway in ART resistance and a Kelch13-dependent mechanism explaining the reduced susceptibility to ART in resistant parasites.

[1]Bernhard Nocht Institute for Tropical Medicine, Bernhard Nocht Str. 74, 20359 Hamburg, Germany. [2]Department of Molecular Biology, Radboud University, Geert Grooteplein 26-28, 6525 GA Nijmegen, Netherlands.
*Present address: Medical Research Council Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge, UK. †These authors contributed equally to this work. ‡Present address: Institute of Physics and Astronomy, Karl-Liebknecht-Straße 24/25, University of Potsdam, 14476 Potsdam, Germany. §Deceased
¶Corresponding author. Email: spielmann@bnitm.de

The Kelch13 compartment and associated proteins defined by a novel BioID approach

We previously found endogenous Kelch13 tagged with green fluorescent protein (GFP) (i.e., expressed from its genomic locus) at foci in close proximity to the parasite's food vacuole (FV) (*17*), a lysosome-like compartment where parasites digest endocytosed host cell cytosol (predominantly hemoglobin). To better define the cellular location of Kelch13, we colocalized it with a series of markers (Fig. 1A and fig. S1). This confirmed the proximity of Kelch13 foci to the FV. Although Kelch13 foci rarely overlapped with secretory pathway markers [endoplasmic reticulum (ER) and Golgi], they were frequently in the vicinity of parts of the ER that were near the FV. Kelch13 foci were also found close to the apicoplast, which itself is located in proximity of the FV. Notably, the resistance-conferring form of Kelch13 (Kelch13^{C580Y}; Cys^{580}→Tyr) fully colocalized with episomally expressed wild-type Kelch13 (wtKelch13) (Fig. 1A, fig. S1, and table S1), indicating that resistance does not involve an altered cellular distribution. Kelch13^{C580Y} also colocalized with episomally expressed hemagglutinin (HA)–tagged Kelch13 (fig. S1). Overall, this analysis suggests that Kelch13 is located at an unknown compartment or cellular structure (henceforth, Kelch13 compartment) not identified by general markers of the secretory system.

To gain insight into the Kelch13 compartment, we used a novel version of BioID (*18*), which we call dimerization-induced quantitative BioID (DiQ-BioID), to identify Kelch13 interactors and compartment neighbors in the living cell. In DiQ-BioID, the biotin ligase BirA* is not directly fused to the target protein but is expressed separately (the biotinylizer) and conditionally recruited to the target (using rapalog), where it biotinylates interaction partners (Fig. 1B). The same parasite culture grown without rapalog (i.e., BirA* not on the target) serves as a control (Fig. 1B), permitting highly specific identification of interactors by quantitative mass spectrometry. We used this approach in parasites expressing FK506 binding protein (FKBP)–tagged Kelch13 from the genomic locus (*17*). Induction of dimerization with rapalog efficiently recruited the biotinylizer to the Kelch13 foci (Fig. 1C), and quantitative mass spectrometry of asynchronous parasites resulted in a list of proteins enriched with very high confidence over control (Fig. 1D, fig. S2, and data S1). Kelch13 was the top hit, confirming that the biotinylizer was successfully recruited to Kelch13, and a band consistent in size with the tagged Kelch13 was detected in anti-biotin blots (fig. S2B). The other hits [designated as Kelch13 interaction candidates (KICs), unless already otherwise named] were predominantly proteins of unknown function. These included proteins that had previously been suspected in ART resistance or were in that context highlighted in genome-wide association studies—for example, an Eps15-like protein (PF3D7_1025000, originally annotated as Formin2) (*19*); ubiquitin carboxyl-terminal hydrolase 1 (UBP1) (*19*–*23*); KIC6 (PF3D7_0609700) (*20*); and MyosinC (MyoC) (*19*) (Fig. 1D, fig. S2, and data S1)—suggesting that the interactome identified a pathway that may be relevant for ART resistance. Modification of the genomic locus to express the endogenous protein fused to GFP (fig. S3) showed that the top hit of these previously highlighted proteins, the Eps15-like protein (henceforth Eps15), colocalized with Kelch13 (Fig. 1E), and coimmunoprecipitation confirmed that Eps15 interacts with Kelch13 (Fig. 1F and fig. S4).

To further validate the Kelch13 DiQ-BioID results and the method per se, we carried out a reverse DiQ-BioID with the endogenously tagged Eps15 (Fig. 1G and fig. S5). This resulted in a list of high-confidence hits that extensively overlapped with the Kelch13 DiQ-BioID (all 16 top hits were also found in the high-confidence hits of the K13 DiQ-BioID), including Kelch13 and UBP1 (Figs. 1G and 2C and data S1). This suggested that the hits are indeed interactors or compartment neighbors of the Kelch13-Eps15 complex and that the DiQ-BioID experiments reached a high depth of coverage. This was confirmed by tagging the corresponding endogenous genes with the sequence encoding GFP (fig. S3), which showed that, besides Eps15, 10 of 12 tested Kelch13 hits colocalize with Kelch13 compartment foci (Fig. 1H, fig. S6, and table S1), demonstrating the power of the method. Only PFK9 and KIC10 did not show an overlap with Kelch13. Of the

10 colocalizing proteins, eight also colocalized in the resistance-relevant ring stage, whereas the other two (KIC8 and KIC9) were not detected in that stage (fig. S6 and table S1). Thus DiQ-BioID identifies interactors and compartment neighbors in living cells with high specificity and in this study defined a series of proteins proximal to the Kelch13-Eps15 complex.

Kelch13-Eps15 marks an AP-2 compartment devoid of clathrin

Because Eps15 in other organisms is typically involved in endocytosis (*24*, *25*), we hypothesized that the Kelch13 compartment might also be involved in this process. In malaria blood stages, endocytosis mediates the large-scale uptake of host cell cytosol (consisting predominantly of hemoglobin), but the molecular details of this process are poorly understood (*26*). In model organisms, the most common and

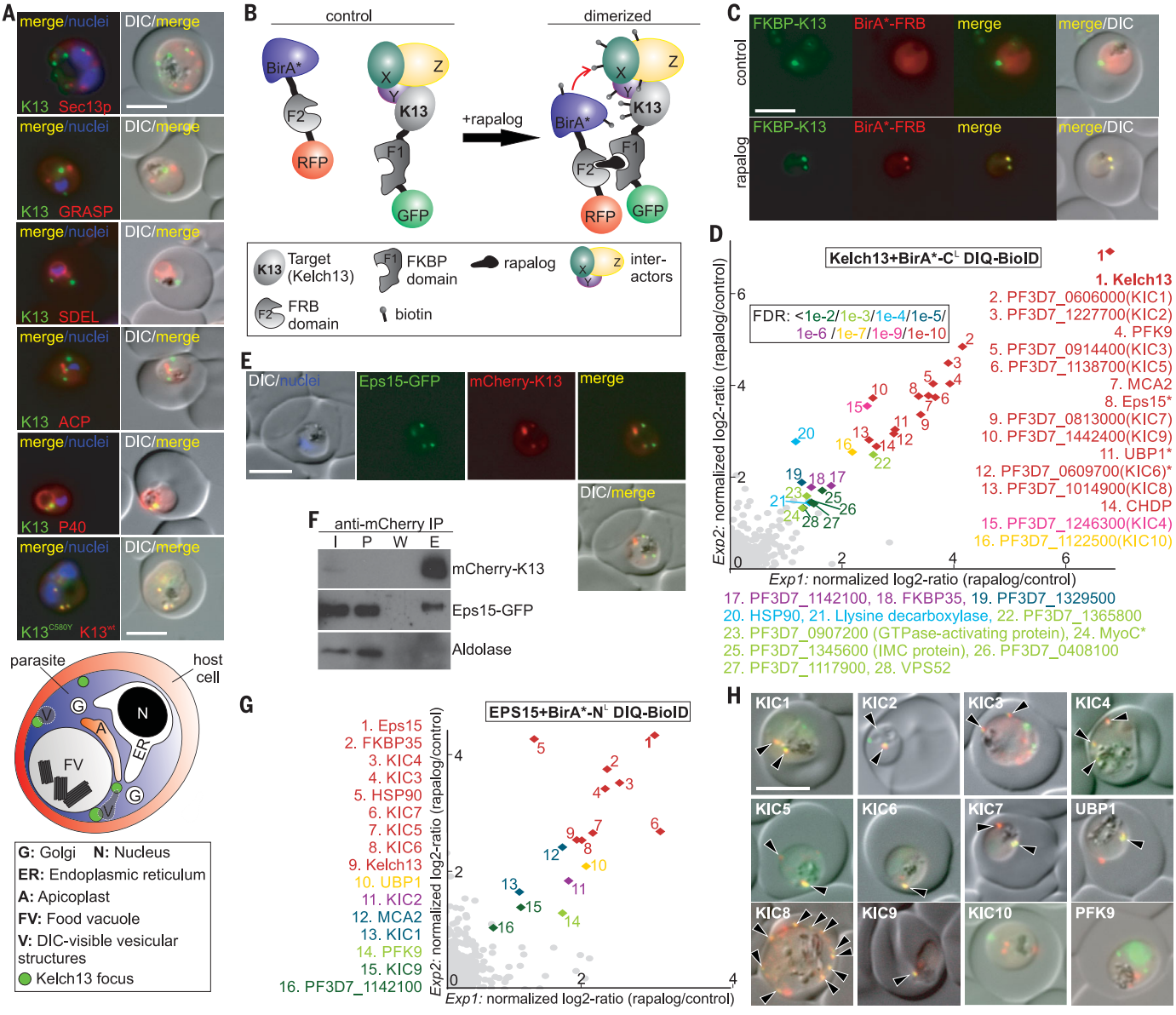


Fig. 1. Kelch13 location and interactome. (**A**) Fluorescence microscopy images showing parasites with endogenously GFP-tagged Kelch13 (K13) (*17*) with coexpressed STVR-SDEL (ER), Sec13p (ER exit sites), P40 (phosphatidylinositol 3-phosphate sensor: food vacuole), ACP (apicoplast), or endogenously GFP-tagged mutated Kelch13 ($K13^{C580Y}$) with episomal mCherry-$Kelch13^{wt}$. See fig. S1 for all panels. (Bottom) Most common arrangement of markers in single-nucleus trophozoites. (**B**) Scheme of DiQ-BioID. (**C**) Rapalog-dependent recruitment of the biotinylizer (BirA*-FRB) to endogenously GFP- and FKBP-tagged Kelch13 (FKBP-K13). (**D**) Top-right quadrant of scatter plot of Kelch13 DiQ-BioID, showing proteins enriched (log_2 ratio) on rapalog (biotinylizer on target) compared with control (biotinylizer cytoplasmic). Significantly enriched proteins are indicated (short, unique names or PlasmoDB identifiers). False discovery rate (FDR) indicated for the least significant experiment. Asterisks indicate proteins previously suspected in ART resistance. Figure S2 shows full plot and replicas with a different biotinylizer. (**E**) Fluorescence microscopy of parasites with endogenously 2xFKBP-GFP-2xFKBP–tagged Eps15 episomally expressing mCherry-Kelch13. (**F**) Immunoprecipitation (IP) of episomally expressed Kelch13 in parasites endogenously expressing GFP-Eps15. I, input extract; P, post IP extract; W, last wash; E, eluate. Full blots and replicas in fig. S4. (**G**) Top-right quadrant of scatter plot of Eps15 DiQ-BioID, as in (D). Figure S5 shows full plot and independent replicas. (**H**) Fluorescence microscopy in parasites with endogenously GFP-tagged DiQ-BioID hits (KIC1-10 or short name) colocalized with episomal mCherry-Kelch13. Arrowheads indicate Kelch13 foci overlapping with the KIC signal (all panels in fig. S6). Scale bars, 5 μm. Merge, merged green and red channel; DAPI, 4′,6-diamidino-2-phenylindole (nuclei).

best studied form of endocytosis is mediated by clathrin, which depends on the AP-2 adaptor complex (*27*, *28*). Although *P. falciparum* Eps15 contains regions specific to malaria parasites, it shows typical hallmarks of Eps15 proteins, including AP-2 binding sites (fig. S5B). To test whether the Kelch13-Eps15 complex marks an endocytic structure, we colocalized Kelch13 with endogenously GFP-tagged clathrin heavy chain (CHC) or AP-2μ (fig. S3). AP-2μ–GFP foci colocalized with (Fig. 2A, arrowhead) or were found in the same region as Kelch13 foci (Fig. 2A, arrow), indicating an endocytosis role for the Kelch13-Eps15 compartment. Unexpectedly, CHC was present in foci that did not overlap with Kelch13 (Fig. 2A). This showed that Kelch13 colocalizes with a compartment that contains AP-2 but is devoid of clathrin. Whereas clathrin typically has additional functions besides endocytosis where it associates with other adaptors, AP-2 is generally associated with clathrin-dependent endocytosis (*27*, *28*). AP-2 independent of clathrin is a highly unusual configuration that, to our knowledge, was so far only observed in *Aspergillus nidulans* (*29*). Of note, AP-2μ has previously been suspected to be involved in ART resistance (*30*, *31*).

To probe this distinction between clathrin and the Kelch13-Eps15 complex and further assess the specificity of the DiQ-BioID procedure, we performed DiQ-BioID with the CHC (fig. S7). This resulted in a list of high-confidence hits differing from the Kelch13 and Eps15 DiQ-BioIDs and included known CHC interactors such as the putative clathrin light chain (even more enriched than the CHC itself) and subunits of the AP-1 adaptor complex (Fig. 2B, fig. S7, and data S1). Unexpectedly, subunits of AP-4—in other organisms considered a clathrin-independent adaptor of the trans-Golgi network (*32*, *33*)—were also enriched, whereas AP-3 was only very mildly enriched, not reaching statistical significance (data S1). AP-4 is not well studied to date, but it is known to interact with tepsin (*34*, *35*). *P. falciparum* tepsin (*34*) was a prominent hit in the CHC DiQ-BioID, validating the AP-4 hits. This suggests that CHC associates with AP-4 in malaria parasites, either indicating differences compared to other organisms or that DiQ-BioID uncovers interactions not retained using previously used approaches in other organisms. Hence, DiQ-BioID reveals a credible CHC interactome that is clearly distinct from the Kelch13-Eps15 interactomes. The CHC DiQ-BioID also detected a different Kelch protein (PF3D7_1205400), potentially indicating that different trafficking complexes harbor distinct Kelch proteins. Only PFK9, HSP90 (heat shock protein 90), and FKBP35 were hits overlapping with the Kelch13 and Eps15 DiQ-BioID experiments, suggesting that these hits might

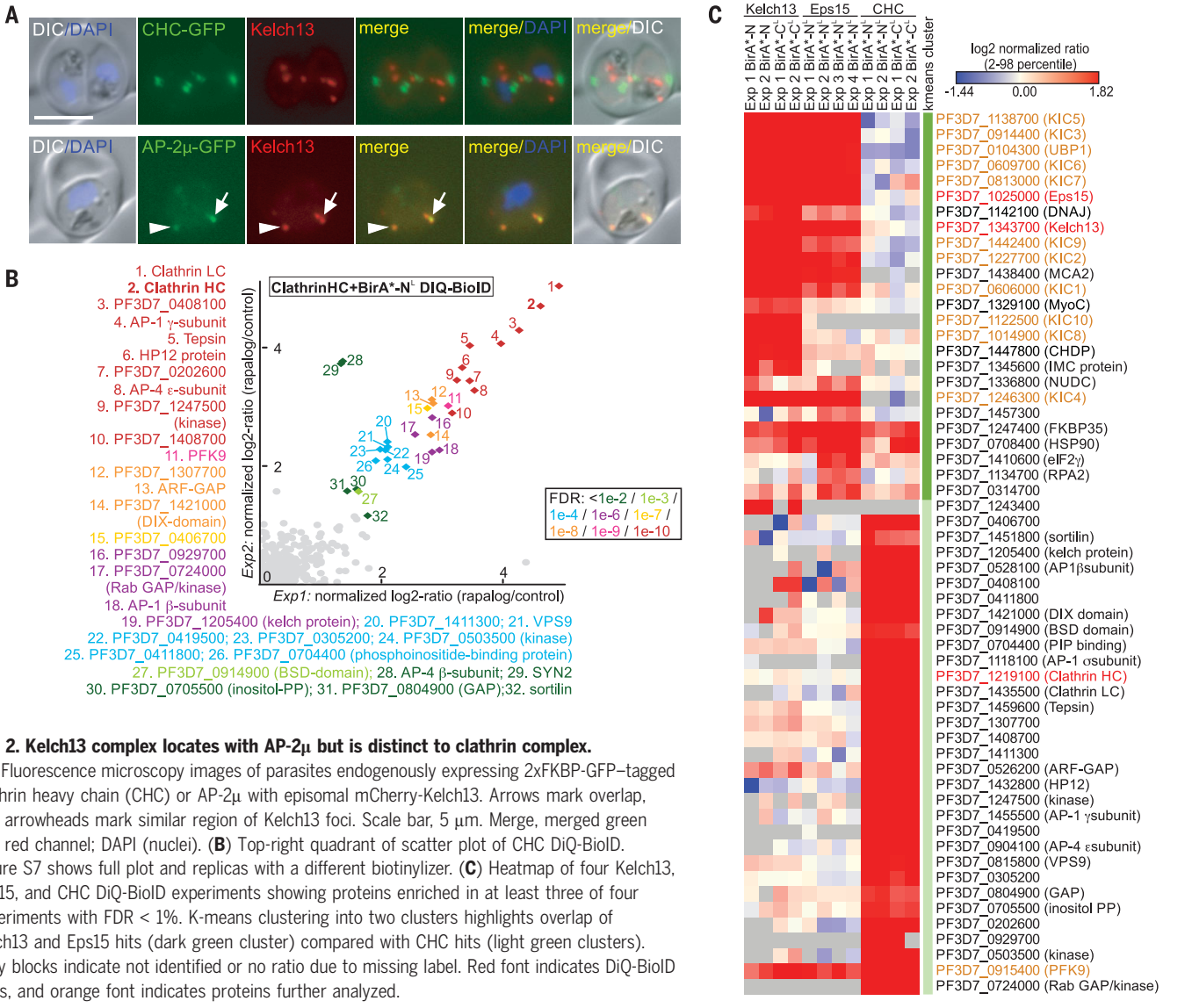


Fig. 2. Kelch13 complex locates with AP-2μ but is distinct to clathrin complex. (**A**) Fluorescence microscopy images of parasites endogenously expressing 2xFKBP-GFP–tagged clathrin heavy chain (CHC) or AP-2μ with episomal mCherry-Kelch13. Arrows mark overlap, and arrowheads mark similar region of Kelch13 foci. Scale bar, 5 μm. Merge, merged green and red channel; DAPI (nuclei). (**B**) Top-right quadrant of scatter plot of CHC DiQ-BioID. Figure S7 shows full plot and replicas with a different biotinylizer. (**C**) Heatmap of four Kelch13, Eps15, and CHC DiQ-BioID experiments showing proteins enriched in at least three of four experiments with FDR < 1%. K-means clustering into two clusters highlights overlap of Kelch13 and Eps15 hits (dark green cluster) compared with CHC hits (light green clusters). Gray blocks indicate not identified or no ratio due to missing label. Red font indicates DiQ-BioID baits, and orange font indicates proteins further analyzed.

result from the method rather than being biologically meaningful (Fig. 2, B and C). This was supported by DiQ-BioID using the biotinylizer alone (fig. S8 and data S2), and a lack of colocalization with Kelch13 was experimentally confirmed in the case of PFK9 (Fig. 1G and fig. S6). Therefore, we did not analyze these proteins further.

Overall, these findings show that Kelch13 defines proteins at a compartment that colocalizes with the classical endocytosis adaptor AP-2μ but that is entirely distinct from clathrin, which typically associates with the AP-2 complex in other organisms. This may indicate the existence of an unusual clathrin-independent endocytosis pathway in malaria parasites.

Kelch13 compartment proteins are involved in endocytosis

To investigate a possible role of the Kelch13-Eps15 compartment in endocytosis, suggested by the presence of Eps15 and colocation with Ap-2μ, we first carried out correlative light and electron microscopy (CLEM) with the parasites expressing endogenously GFP-tagged Eps15. This revealed that Eps15 resides in proximity of host cell cytosol-filled membranous structures in the parasite in three of three analyzed cells and in five of five foci in these cells (Fig. 3A and fig. S9, A to C). Hence, it can be inferred that Kelch13, Eps15, AP-2, UBP1, and the nine colocalizing KICs are present proximal to host cell cytosol-filled structures in the parasite, congruent with a role of the Eps15-Kelch13 complex in endocytosis of host cell cytosol. In support of this, we noted that in >90% of cells with vesicle-like structures visible in differential interference contrast (DIC), Kelch13 foci were found at such structures (fig. S9D), suggesting that these may represent the host cell cytosol-filled structures observed by CLEM and likely are cytostomes.

To more directly assess the function of KICs, we first carried out a selection-linked integration gene disruption (SLI-TGD) screen (*17*) to establish which of the Kelch13 compartment proteins are essential for parasite survival. Of the 13 tested genes, only the ones encoding Eps15, UBP1, and KIC7 were refractory to disruption and hence are likely essential for parasite survival (Fig. 3B and fig. S3). Knock sideways (KS) (*17*) to conditionally inactivate these proteins, as well as AP-2μ, confirmed their importance for parasite growth (Fig. 3, B and C, and fig. S10). Note that the smaller effect on growth after inactivating Eps15 likely is the result of inefficient KS (Fig. 3C and fig. S10). To test whether Kelch13, UBP1, KIC7, Eps15, and AP-2μ are involved in endocytosis, we conditionally inactivated them by KS (fig. S10) and assessed hemoglobin uptake into the parasite over 8 hours using a previously established assay, which results in a bloated FV phenotype

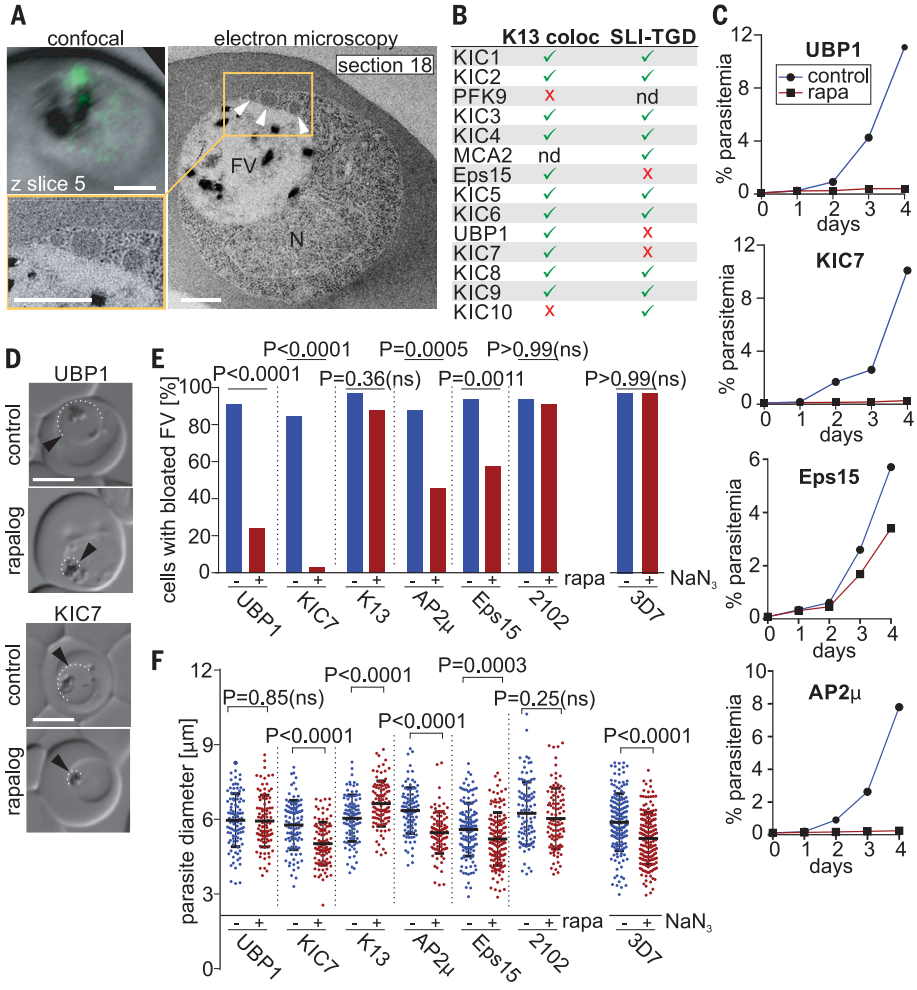


Fig. 3. Kelch13 compartment proteins are involved in endocytosis of host cell cytosol. (**A**) CLEM with a cell expressing endogenously tagged Eps15-GFP: confocal image (scale bar, 2 μm), EM section and magnification thereof (scale bars, 0.5 μm). Arrowheads indicate vesicles containing material of host cell density. N, nucleus; FV, food vacuole. Details and two more cells in fig. S9. (**B**) Summary of Kelch13 colocalization and SLI-TGD screen of DiQ-BioID hits. nd, not done. (**C**) Growth curves of parasites after induction of KS. One of three independent experiments. Figure S10 shows replicas and details on KS. (**D**) Example DIC images showing bloated (control) and nonbloated (inhibited endocytosis) FVs (partly highlighted: dashed line and arrows) after KS (rapalog) and E64 treatment. (**E**) Quantification of number of cells with bloated FVs after E64 treatment for control (no rapalog) and KS (rapa) for 8 hours (*n* for control/rapa: UPB1, 33/33; KIC7, 33/33; Kelch13, 32/33; AP-2μ, 33/32; Eps15, 33/33; 2102, 33/33; 3D7 NaN_3, 34/34). Fisher's exact test. Replicas in fig. S10. (**F**) Growth (cell diameter) of cells scored for bloated FVs after KS compared with control, data pooled from all three (five for NaN3) independent experiments, with a total of *n* = 99/97 cells for no rapalog/rapa with UBP1, 101/97 with KIC7, 134/134 with Eps15, 99/97 with AP-2μ, 100/100 with 2102, and 170/170 for 3D7 ± NaN_3. Unpaired, two-tailed *t* test. ns, not significant.

if endocytosis is operational (*36*). Inactivation of UBP1, KIC7, Eps15, and AP-2μ significantly reduced transport of hemoglobin to the FV, whereas the controls or inactivation of Kelch13 or an unrelated essential protein [PF3D7_0210200 (*17*), here called 2102] did not (Fig. 3D). Inactivation of KIC7, AP-2μ, and to a lesser extent Eps15, reduced growth over the 8-hour assay time. However, the growth defect itself was not the reason for the phenotype, as inducing an unrelated growth defect using azide did not impair endocytosis (Fig. 3E). Surprisingly, inactivation of Kelch13 significantly increased parasite size compared with the control, and inactivation of UBP1 showed a clear reduction of endocytosis but no significant reduction in cell size over the 8-hour assay time (Fig. 3, D to F). We conclude that proteins of the Kelch13 compartment play a role in endocytic uptake of hemoglobin, but Kelch13 itself does not appear to be necessary for this process in trophozoites. The lack of

effect in trophozoites agrees with the previous finding that inactivation of Kelch13 specifically affected the ring stage only (*17*).

A detailed comparison of the stage-specific essentiality of Kelch13, UBP1, and KIC7 revealed that all three proteins caused a similar phenotype when inactivated in rings (severe delay of ring stage with limited progression to the trophozoite stage). In contrast, inactivation in early trophozoites prevented development into schizonts in KIC7 and UBP1, but not in Kelch13 (fig. S11). This suggests a similar function of these proteins but that Kelch13 is not essential for this function in trophozoites. A previously established KS cell line that only partially inactivates Kelch13 (Kelch13 3xNLS) and that does not noticeably affect parasite growth in a 4-day growth assay (*17*) showed a mild delay of the ring stage (fig. S11). Hence, while not lethal, partial inactivation of Kelch13 (3xNLS line) leads to a much milder, but similar, effect compared with full inactivation of Kelch13 (1xNLS line).

Kelch13 influences endocytosis and resistance in ring stages

Hemoglobin by-products activate ART, and decreased hemoglobin digestion reduces ART susceptibility (*37*–*39*). Hence, our data might indicate that reduced levels of endocytic uptake of host cell cytosol could, by controlling the supply of hemoglobin available for digestion and ART activation, be the mechanism of ART resistance. However, this would need to take place in the resistance-relevant ring stage (*4*). As UBP1 and KIC7 affected endocytosis in trophozoites and showed a similar phenotype in rings to Kelch13 (fig. S11), we reasoned that Kelch13 might influence endocytosis specifically in ring stages. However, as it is not clear whether young ring stages endocytose hemoglobin at all (*39*–*41*), we decided to first clarify this. Because of the overabundance of hemoglobin in the parasite culture, we used host cells loaded with fluorescent dextrans, a well-established tool for studying endocytic uptake in malaria parasites (*36*, *40*). These experiments demonstrated uptake of host cell material into young ring stages (Fig. 4A).

Next, we tested the role of Kelch13 in endocytosis in 0-to-6-hour young rings [the time the drug pulse is applied in ART susceptibility assays (*4*)] using the Kelch13 3xNLS parasites (the cell line leading to only partial inactivation of Kelch13, to avoid detrimental effects on parasite growth). Inactivation of Kelch13 significantly reduced endocytic uptake in these rings (Fig. 4B and fig. S12). Inactivation of KIC7 similarly reduced endocytosis in rings (Fig. 4C and fig. S12), congruent with its similar KS phenotype to Kelch13 in rings and its role in endocytosis in trophozoites (Fig. 3E). Hence, KIC7 affects endocytosis in both rings and trophozoites and, together with the similar stage-specific phenotypes, suggests that the Kelch13 compartment proteins are needed for endocytosis in all asexual blood stages, with the exception of Kelch13, which is only needed for this process in rings.

To assess whether this could be the reason for resistance, we next tested hemoglobin uptake into rings of resistant parasites containing a mutated genomic Kelch13^{C580Y} (*17*). The resistant parasites also displayed reduced endocytic uptake (Fig. 4D and fig. S12), suggesting this as a reason for reduced susceptibility to ART. To further confirm this hypothesis, we inactivated Kelch13 (again using the Kelch13 3xNLS line) and measured whether this can render parasites ART-resistant using a standard RSA (*4*). When Kelch13 was inactivated, parasites showed levels of resistance comparable to the resistant Kelch13^{C580Y} parasite line (Fig. 4E). Thus, Kelch13 inactivation reduces endocytosis of host cell material and increases resistance. We conclude that early ring stages already endocytose hemoglobin, and that this process is reduced in parasites with a resistance-conferring mutation in Kelch13 or in which Kelch13 is inactivated, and that this correlates with ART resistance.

Kelch13 compartment proteins are involved in ART resistance

Having established that inactivation of Kelch13 induces ART resistance, we next assessed whether other Kelch13 compartment proteins also play a role in ART resistance. First, we used the cell lines with the disrupted, nonessential KICs. Whereas parasites with disrupted KIC1, 2, 3, 6, 8, and 9, as well as KIC10 (which is not Kelch13 compartment–associated), did not show an altered response to ART, disruption of KIC4, KIC5, and MCA-2 led to a reduced susceptibility in a standard RSA (Fig. 4F). While the reduced susceptibility correlated with a low growth rate of the KIC5 and MCA-2 disruption lines (suggesting some importance of these proteins for efficient parasite growth), there was little correlation overall between growth and ART resistance (Fig. 4G), excluding growth levels as a factor in resistance. For instance, the line with a disruption of KIC10, the only non–Kelch13 compartment KIC (Fig. 1H), had the third lowest growth rate but did not show resistance (Fig. 4, F and G).

To assess the role of the essential proteins AP-2μ, Eps15, UBP1, and KIC7 in ART resistance, we partially inactivated them (using KS) before and during the ART drug pulse of the RSA in a manner that did not kill the parasites. Inactivation of KIC7, AP-2μ, Eps15, and UBP1 induced ART resistance (Fig. 4H). In contrast, reducing parasite viability with azide or inactivation of an unrelated essential protein [2102 (*17*)] did not affect the outcome of the RSA, indicating that the effect on ART susceptibility is specific for the function of Kelch13 and its compartment proteins (Fig. 4H). We conclude that inactivation of more than half (7 of 13) of the tested Kelch13 compartment proteins, as well as Kelch13 itself, reduces the responsiveness to ART. Notably, all of the Kelch13 compartment proteins affecting endocytosis (Fig. 3, D and E) also rendered parasites resistant, establishing a strong link between hemoglobin uptake and ART resistance (table S1). We conclude that Kelch13 compartment proteins are part of an ART resistance pathway, and the correlation of reduced endocytosis and resistance indicates that this is the mechanism of this mode of resistance. To test whether the corresponding genes belong to a coregulated pathway, we assessed their stage-specific transcription patterns across asexual blood stage development (*42*). Except for *mca2*, all genes involved in resistance clustered into one group when compared with (i) the other validated Kelch13 compartment hits (Fig. 4I), (ii) all significant Kelch13 interactome hits (fig. S13A), and (iii) all hits significant in any of the BioIDs (fig. S13B). In contrast, the other hits showed widely differing transcription patterns, reflecting the asynchronous nature of the parasites used as input for the DiQ-BioID. Overall, the coexpression lends independent support to the idea that the genes experimentally identified here belong to a coregulated pathway involved in hemoglobin endocytosis.

To show that the resistance pathway of the Kelch13 interactors is clinically relevant, we chose a mutation in UBP1 (R3138H) that was identified as a possible contributor to ART resistance by genomic surveillance of parasite field samples (*19*). We changed this position in UBP1 in 3D7 parasites and found that this rendered these parasites resistant to ART (Fig. 4J) at a level comparable to the resistance demonstrated when UBP1 was inactivated by KS (Fig. 4H). These findings demonstrate that the pathway identified here can be clinically relevant and that mutations in the UBP1 can modulate ART resistance.

Reduced levels of Kelch13 explain ART resistance

To determine how mutated Kelch13 influences resistance, we first assessed whether the interaction profile of mutated Kelch13 was altered compared with wtKelch13. However, DiQ-BioID with the Kelch13^{C580Y} parasite line revealed no marked differences to the DiQ-BioID with wtKelch13 (Fig. 5A, fig. S14, and table S1). The only exception was KIC10, which was less enriched in the Kelch13^{C580Y} DiQ-BioID, but this is the only KIC not colocalizing with Kelch13 and has no role in resistance (Figs. 1H and 4F and table S1). As these findings indicated that there is no change in the interaction profile of Kelch13^{C580Y} compared with wtKelch13, we hypothesized that mutating Kelch13 does not change a specific quality of its function but

reduces its overall activity. This idea is also suggested by our finding that reduced activity of Kelch13 causes resistance (Fig. 4E). In addition, previous work indicated that resistant parasites harbor less Kelch13 protein (*43*). In agreement with this observation, we found ~30 to 50% less Kelch13 in parasites with the resistance-conferring form in Western blots and by measuring Kelch13 levels directly in living cells (Fig. 5, B to D). To test whether increased expression of Kelch13 can lead to the loss of resistance, we episomally expressed different versions of Kelch13 on the resistant genomic *kelch13*^{C580Y} background. Adding wtKelch13 (leading to expression of episomal

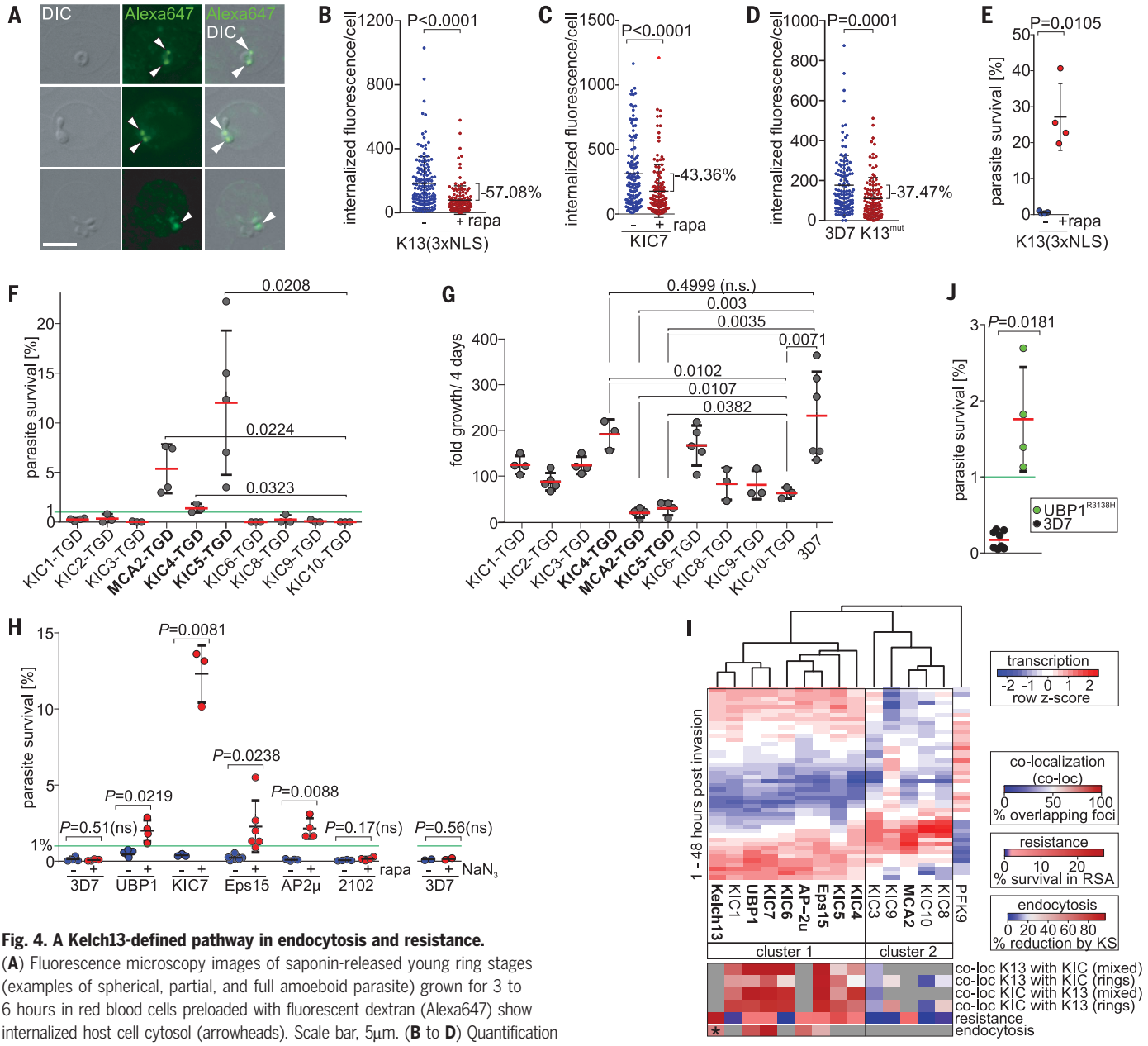


Fig. 4. A Kelch13-defined pathway in endocytosis and resistance.
(**A**) Fluorescence microscopy images of saponin-released young ring stages (examples of spherical, partial, and full amoeboid parasite) grown for 3 to 6 hours in red blood cells preloaded with fluorescent dextran (Alexa647) show internalized host cell cytosol (arrowheads). Scale bar, 5µm. (**B** to **D**) Quantification of dextran uptake into 3-to-6-hour rings after inactivation (rapa) of Kelch13 by KS in the 3xNLS line (*17*) (B) or KIC7 (C) compared with control (no rapa) or in the K13^{mut} parasites compared with wild type (3D7) (D). Each point shows the amount of fluorescence internalized into one cell (arbitrary fluorescence units). Data pooled from three independent experiments (individual experiments in fig. S12) with a total of *n* = 138, 140, 132, 130, 122, and 114 cells for Kelch13 KS control, rapa, KIC7 control, KIC7 rapa, 3D7, and K13^{mut}, respectively. Unpaired, two-tailed *t* test. *P* values and percent reduction of the mean are indicated. (**E**) RSAs after inactivating Kelch13 by KS (rapa) compared with control with Kelch13 3xNLS parasites (*17*). (**F**) RSAs with the TGD lines indicated. (**G**) Growth rate of TGD lines. (**H**) RSAs after KS (rapa) to inactivate the proteins indicated compared with control or in 3D7 after inhibiting growth using NaN_3. (**I**) Heatmap of transcription levels in 1-hour intervals across asexual blood stage development [values from (*42*)], clustered according to similarity for the proteins analyzed in this study (see also fig. S13). Note that KIC2 was not part of the dataset. Bottom shows data of colocalization, RSA, and endocytosis experiments (table S1; Fig. 4, B, F, and H; and Fig. 3E; respectively) as heatmaps (color shading indicated; gray, not applicable). Asterisk indicates ring-stage endocytosis assay. (**J**) RSA with parasites harboring a mutated *ubp1* gene encoding a change from Arg to His at amino acid position 3138 (*19*). Resistance in RSA defined as mean survival above 1% [green line in (F), (H), and (J)], a previously set standard (*4*). [(B) to (H), and (J)] Error bars show SD (not shown if close to mean). *P* values indicated [only selected *P* values in (F) and (G)]; ns, not significant. [(E) to (H), and (J)] Unpaired, two-tailed *t* test with Welch's correction. Each point indicates an independent experiment.

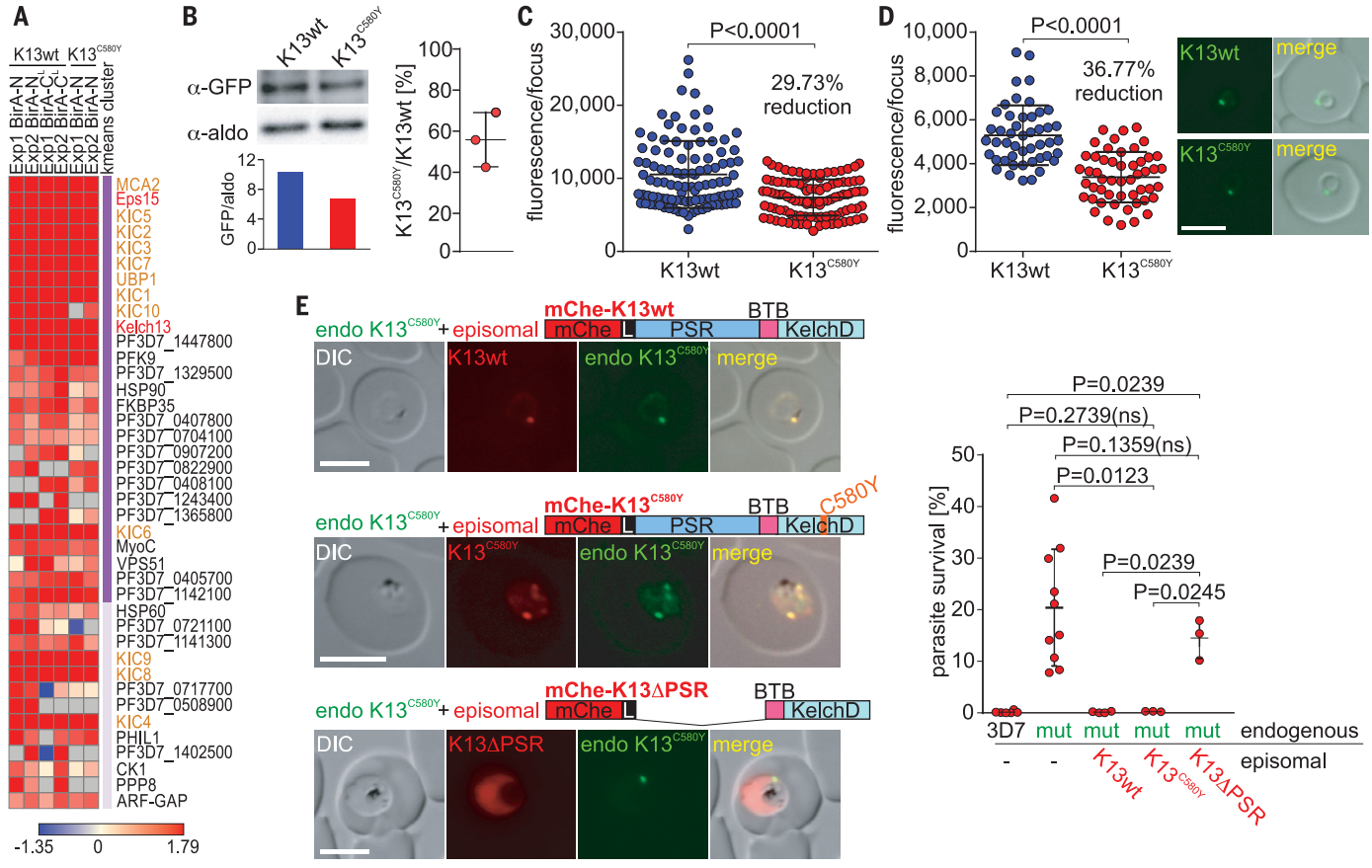


Fig. 5. Reduced Kelch13 activity leads to ART resistance. (**A**) Heatmap of DIQ-BioID data of the four Kelch13 and the two Kelch13^{C580Y} experiments [hits with FDR < 1% in at least two of four (Kelch13) or two (Kelch13^{C580Y}) experiments]. Color intensities: 2nd to 98th percentile of $\log_2$ rapalog/control normalized ratios of all proteins in this analysis. NaN (data S3) values in gray. Rows were clustered and gene identifiers were color coded as in Fig. 2C. (**B**) Immunoblot-based quantification of wtKelch13 (K13wt) and Kelch13^{C580Y} (K13^{C580Y}) levels in parasites expressing the otherwise-identical GFP-fused protein from the endogenous locus. (Left) An example band pair with Kelch13 levels relative to the loading control (aldolase) shown as bars below the blot (axis shows intensity ratio). (Right) K13^{C580Y} levels compared with K13wt (n = 3 independent experiments; each experiment is shown by a point that derives from the average of a serial dilution of K13wt and K13^{C580Y} run on the same gel and normalized relative to the respective aldolase control). (**C** and **D**) Quantification of Kelch13 foci fluorescence intensity in trophozoites (n = 105, 102 foci) and rings (n = 47, 51 foci) in K13wt and K13^{C580Y} parasites, respectively. Data pooled from three independent experiments; percent reduction of the mean is indicated. Example fluorescence images of ring stages are shown. (**E**) Fluorescence microscopy images of parasites with endogenous Kelch13^{C580Y} episomally expressing the construct indicated (PSR, plasmodium-specific region; L, linker) and RSA with the respective cell lines (right). Cell lines expressing mCherry-K13wt and mCherry-K13^{C580Y} were imaged after smearing, mCherry-K13ΔPSR live. Size bar, 5 μm. (**F**) Model of resistance and summary of results of this study. Asterisk indicates only in ring stages; nd, not determined (see also table S1). Bars show SD, mean is indicated. Unpaired, two-tailed t test with Welch's correction. P values are indicated [in (E), only for selected pairs]; ns, not significant.

wtKelch13 and endogenous Kelch13^{C580Y}) rendered these resistant parasites fully sensitive to dihydroartemisinin (DHA) again (Fig. 5E). To test whether this was due to an additional property held by wtKelch13 that is lacking in Kelch13^{C580Y}, we episomally expressed Kelch13^{C580Y} on the resistant background (resulting in parasites expressing episomal Kelch13^{C580Y} and endogenous Kelch13^{C580Y}, in effect only raising the abundance of Kelch13^{C580Y}). This also reverted the parasites from resistant to sensitive again (Fig. 5E), demonstrating that reduced protein levels (or reduced activity) alone explains resistance, congruent with the finding that partially inactivating Kelch13 has a similar effect (Fig. 4E). In contrast, Kelch13 missing the *Plasmodium*-specific region (PSR), a modification that led to a loss of the correct location of this construct (Fig. 5E), did not revert the resistance phenotype when added on top of the *kelch13^{C580Y}* background (Fig. 5E). This demonstrated the specificity of the effect with episomal Kelch13^{C580Y} and indicated that the BTB and Kelch13 domain without the PSR are not functional.

Overall, these results show that reduced activity of Kelch13 is the most likely mechanism of how Kelch13 mutation causes resistance and that this reduced activity, at least in part, is due to decreased Kelch13 levels in resistant parasites. A generally reduced function also explains why inactivation of several Kelch13 compartment proteins renders parasites resistant, as each will reduce the output of a common pathway that, based on our functional data, is endocytosis.

Discussion

Here, we show that Kelch13 defines an endocytosis pathway required for the uptake of host cell hemoglobin and that this pathway is critical for ART resistance. Although it is well established that ARTs are activated by the parasite's digestion products of hemoglobin and that this is a prerequisite for ART action (*37*–*39*), this process was not considered to be involved in resistance in parasites with Kelch13 mutations (*15*, *16*). Instead, potential roles of Kelch13 downstream of ART activation, e.g., in mitigating the ART-induced cellular stress response, are the current target of interest and central to hypotheses on the mechanism of resistance (*15*, *16*). However, our data indicate that Kelch13 and its compartment proteins mediate resistance upstream of both, drug activation and action. We propose a model where Kelch13 and its compartment proteins control endocytosis levels, thereby influencing the amount of hemoglobin available for degradation and hence the concentration of active drug (Fig. 5F). This mechanism explains the slowed development of ring stages observed in ART-resistant parasites (*9*, *44*), as the reduction in hemoglobin endocytosis in rings diminishes the supply of amino acids. This agrees with the finding that ART-resistant parasites depend more on exogenous amino acids to mature from rings to trophozoites than wild-type parasites (*45*) and that removing functions needed for amino acid access of the parasite prolongs the ring stage (*46*). Reduced levels of amino acids could also account for some of the changes observed in ART-resistant parasites, such as elevated cellular stress responses (*9*, *15*). It also indicates that the fitness cost incurred by resistance-conferring Kelch13 mutations (*47*, *48*) is a direct result of the resistance mechanism. Hence, there is a trade-off between resistance and growth levels in ring stages (Fig. 5F), creating a pressure for compensatory adaptations that may explain the importance of the genetic background in parasite resistance and fitness (*8*, *48*).

Altered endocytosis as a mechanism of ART resistance reconciles aspects of previous findings relating to ART resistance, e.g., a role of phosphatidylinositol 3-phosphate (and its generating kinase) (*12*), a membrane signature specific for endosomes (*49*) that is also present on the endolysosomal system of malaria parasites (*36*, *50*), or of coronin (*51*) that could influence endocytosis via actin. We also validated the suspected role of AP-2μ in resistance (*30*, *31*) and show that, as in other organisms (*24*), this protein is involved in endocytosis. This indicates that Kelch13 marks a clathrin-independent endocytosis pathway that unexpectedly still contains the typical clathrin adaptor AP-2μ. Besides Eps15 and AP-2μ, KIC4 shows homology to alpha adaptins [according to HHPred (*52*)], adding an additional protein typical for endocytosis in a generally highly derived pathway.

In contrast to its interactors, inactivation of Kelch13 impaired endocytosis only in rings, not in trophozoites. This likely is the reason why Kelch13, among all the proteins influencing endocytosis, is the one frequently found mutated in resistant parasite isolates (*53*). Complete inactivation of Kelch13 arrests growth in ring stages (*17*), and a similar phenotype was here observed with KIC7 and UBP1. The stage-specific effects of inactivating Kelch13 are observed in a milder form in ART-resistant parasites that display a prolonged ring stage followed by an accelerated development in later stages (*9*, *44*), and this observation was here recapitulated by partial inactivation of Kelch13. This indicates that resistance-conferring mutations partially reduce Kelch13 function, a hypothesis supported by our finding that partial inactivation of Kelch13 in rings induces parasite resistance and that additional expression of Kelch13^{C580Y} on a resistant background with the same mutation is sufficient to render parasites fully sensitive again. As resistant parasite field isolates harbor less Kelch13 (*43*), which we also found in our resistant laboratory line here, but transcript levels appear to be unchanged (*9*), it can be assumed that resistance-conferring mutations alter Kelch13 protein stability.

We envisage that the mechanism of ART resistance indicated by this work will aid in finding ways to antagonize it. It may also inform the choice of ART partner drugs, particularly as hemoglobin digestive processes are the target of existing drugs. Finally, the here-identified proteins of the resistance pathway are candidates for novel parasite markers influencing ART resistance that now can be assessed in population studies, as illustrated by the identification of a resistance-conferring mutation in UBP1. It should, however, be noted that resistance-causing mutations are most likely to occur in proteins that have the least impact on trophozoite survival, such as Kelch13.

REFERENCES AND NOTES

1. WHO, "Guidelines for the treatment of malaria" (World Health Organization, ed. 3, 2015).
2. A. M. Dondorp *et al.*, *N. Engl. J. Med.* **365**, 1073–1075 (2011).
3. A. M. Dondorp *et al.*, *N. Engl. J. Med.* **361**, 455–467 (2009).
4. B. Witkowski *et al.*, *Lancet Infect. Dis.* **13**, 1043–1049 (2013).
5. N. Klonis *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5157–5162 (2013).
6. S. Saralamba *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 397–402 (2011).
7. F. Ariey *et al.*, *Nature* **505**, 50–55 (2014).
8. J. Straimer *et al.*, *Science* **347**, 428–431 (2015).
9. S. Mok *et al.*, *Science* **347**, 431–435 (2015).
10. F. Rocamora *et al.*, *PLOS Pathog.* **14**, e1006930 (2018).
11. J. Gibbons *et al.*, *BMC Genomics* **19**, 849 (2018).
12. A. Mbengue *et al.*, *Nature* **520**, 683–687 (2015).
13. C. Dogovski *et al.*, *PLOS Biol.* **13**, e1002132 (2015).
14. M. Zhang *et al.*, *Cell Host Microbe* **22**, 766–776.e4 (2017).
15. L. Tilley, J. Straimer, N. F. Gnädig, S. A. Ralph, D. A. Fidock, *Trends Parasitol.* **32**, 682–696 (2016).
16. K. Haldar, S. Bhattacharjee, I. Safeukui, *Nat. Rev. Microbiol.* **16**, 156–170 (2018).
17. J. Birnbaum *et al.*, *Nat. Methods* **14**, 450–456 (2017).
18. K. J. Roux, D. I. Kim, M. Raida, B. Burke, *J. Cell Biol.* **196**, 801–810 (2012).
19. G. C. Cerqueira *et al.*, *Genome Biol.* **18**, 78 (2017).
20. P. Hunt *et al.*, *Mol. Microbiol.* **65**, 27–40 (2007).
21. P. Hunt *et al.*, *BMC Genomics* **11**, 499 (2010).
22. S. Borrmann *et al.*, *Sci. Rep.* **3**, 3318 (2013).
23. G. Henriques *et al.*, *J. Infect. Dis.* **210**, 2001–2008 (2014).
24. H. T. McMahon, E. Boucrot, *Nat. Rev. Mol. Cell Biol.* **12**, 517–533 (2011).
25. F. Tebar, T. Sorkina, A. Sorkin, M. Ericsson, T. Kirchhausen, *J. Biol. Chem.* **271**, 28727–28730 (1996).
26. S. E. Francis, D. J. Sullivan Jr., D. E. Goldberg, *Annu. Rev. Microbiol.* **51**, 97–123 (1997).
27. Z. Kadlecova *et al.*, *J. Cell Biol.* **216**, 167–179 (2017).
28. M. Kaksonen, A. Roux, *Nat. Rev. Mol. Cell Biol.* **19**, 313–326 (2018).
29. O. Martzoukou, S. Amillis, A. Zervakou, S. Christoforidis, G. Diallinas, *eLife* **6**, e20083 (2017).
30. G. Henriques *et al.*, *Malar. J.* **12**, 118 (2013).
31. G. Henriques *et al.*, *Antimicrob. Agents Chemother.* **59**, 2540–2547 (2015).
32. J. Hirst, N. A. Bright, B. Rous, M. S. Robinson, *Mol. Biol. Cell* **10**, 2787–2802 (1999).
33. E. C. Dell'Angelica, C. Mullins, J. S. Bonifacino, *J. Biol. Chem.* **274**, 7278–7285 (1999).
34. G. H. Borner *et al.*, *J. Cell Biol.* **197**, 141–160 (2012).
35. A. K. Davies *et al.*, *Nat. Commun.* **9**, 3958 (2018).
36. E. Jonscher *et al.*, *Cell Host Microbe* **25**, 166–173.e5 (2019).
37. N. Klonis, D. J. Creek, L. Tilley, *Curr. Opin. Microbiol.* **16**, 722–727 (2013).
38. N. Klonis *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 11405–11410 (2011).
39. S. C. Xie *et al.*, *J. Cell Sci.* **129**, 406–416 (2016).

40. N. Abu Bakar, N. Klonis, E. Hanssen, C. Chan, L. Tilley, *J. Cell Sci.* **123**, 441–450 (2010).
41. L. H. Bannister, J. M. Hopkins, G. Margos, A. R. Dluzewski, G. H. Mitchell, *Microsc. Microanal.* **10**, 551–562 (2004).
42. Z. Bozdech *et al*., *PLOS Biol.* **1**, e5 (2003).
43. G. Siddiqui, A. Srivastava, A. S. Russell, D. J. Creek, *J. Infect. Dis.* **215**, 1435–1444 (2017).
44. A. Hott *et al*., *Antimicrob. Agents Chemother.* **59**, 3156–3167 (2015).
45. D. Bunditvorapoom *et al*., *Sci. Rep.* **8**, 12622 (2018).
46. P. Mesén-Ramírez *et al*., *PLOS Biol.* **17**, e3000473 (2019).
47. S. Nair *et al*., *Antimicrob. Agents Chemother.* **62**, e00605-18 (2018).
48. J. Straimer *et al*., *mBio* **8**, e00172-17 (2017).
49. T. Balla, *Physiol. Rev.* **93**, 1019–1137 (2013).
50. L. Tawk *et al*., *Eukaryot. Cell* **9**, 1519–1530 (2010).
51. A. R. Demas *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12799–12804 (2018).
52. L. Zimmermann *et al*., *J. Mol. Biol.* **430**, 2237–2243 (2018).
53. S. Takala-Harrison *et al*., *J. Infect. Dis.* **211**, 670–679 (2015).

ACKNOWLEDGMENTS

This article is dedicated to Hendrik Herrmann. We thank L. Tilley for suggestions of the timing of Kelch13 KS induction for RSAs. We thank Jacobus Pharmaceuticals for supplying WR99210. DSM1 (MRA-1161) was received from MR4/BEI Resources, NIAID, NIH. **Funding:** E.J., S.F., A.B.S., and T.S. acknowledge funding by the Research Training Group (GRK 1459) of the German Research Foundation (DFG). S.Schm. acknowledges funding by the German Center for Infection Research (TTU 03.806), M.S. thanks the Jürgen Manchot Stiftung and S.Scha. the Claussen-Simon Stiftung for funding. W.A.M.H. obtained financial support from Netherlands Organization for Scientific Research (NWO-VENI, 722.012.004). **Author contributions:** J.B. designed research, carried out experiments, analyzed data, prepared figures, and wrote the methods section. S.Scha., S.Schm., E.J., and S.F. designed and carried out experiments, analyzed data, and prepared figures. M.S., R.S., B.B., U.F., P.M.-R., and A.B.S. carried out experiments. W.A.M.H. and C.G.T. conducted the mass spectrometry experiments, analyzed data, and prepared figures. R.B. supervised the mass spectrometry experiments and analyzed data. T.S. conceived of and supervised the project, designed experiments, carried out imaging and image analysis, prepared figures, and wrote the paper with critical input from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials, and the Kelch13, Eps15, and CHC mass spectrometry data are available via ProteomeXchange with identifier PXD008219.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6473/51/suppl/DC1
Materials and Methods
Fig. S1 to S14
Table S1
References (*54*–*78*)
Data S1 to S4

View/request a protocol for this paper from *Bio-protocol*.

26 March 2019; resubmitted 19 September 2019
Accepted 30 October 2019
10.1126/science.aax4735

REPORT

TOPOLOGICAL OPTICS

A single photonic cavity with two independent physical synthetic dimensions

Avik Dutt[1], Qian Lin[2], Luqi Yuan[3]*, Momchil Minkov[1], Meng Xiao[4], Shanhui Fan[1]*

The concept of synthetic dimensions has generated interest in many branches of science, ranging from ultracold atomic physics to photonics, as it provides a versatile platform for realizing effective gauge potentials and topological physics. Previous experiments have augmented the real-space dimensionality by one additional physical synthetic dimension. In this study, we endow a single ring resonator with two independent physical synthetic dimensions. Our system consists of a temporally modulated ring resonator with spatial coupling between the clockwise and counterclockwise modes, creating a synthetic Hall ladder along the frequency and pseudospin degrees of freedom for photons propagating in the ring. We observe a wide variety of physics, including effective spin-orbit coupling, magnetic fields, spin-momentum locking, a Meissner-to-vortex phase transition, and signatures of topological chiral one-way edge currents, completely in synthetic dimensions. Our experiments demonstrate that higher-dimensional physics can be studied in simple systems by leveraging the concept of multiple simultaneous synthetic dimensions.

There has been interest in creating synthetic dimensions to study classical and quantum dynamics (*1*) in systems with extra dimensions beyond their real-space geometric dimensionality (*2*). Synthetic dimensions can be formed by coupling atomic or photonic states with different internal degrees of freedom to form a lattice. These degrees of freedom may be based on the frequency, spin, linear momentum, orbital angular momentum, spatial supermodes, or arrival time of light pulses (*3*). Previous experiments have provided demonstrations of $(d + 1)$-dimensional physics on d-dimensional real-space lattices by using one extra synthetic dimension, for $d = 1$ (*4*, *5*) or $d = 0$ (*6*–*8*). Although theoretical proposals exist for creating two or more separate synthetic dimensions (*9*, *10*), these proposed phenomena have thus far eluded experimental observation. The realization of two or more synthetic dimensions markedly simplifies the experimental requirements for studying a rich set of topologically nontrivial phenomena—e.g., the high-dimensional quantum Hall effect (*11*, *12*)—without the need for complex higher-dimensional structures in real space.

We demonstrate a system exhibiting two independent physical synthetic dimensions. Our system (Fig. 1A) consists of a ring resonator supporting a synthetic frequency dimension formed by the longitudinal cavity modes, as well as a synthetic pseudospin dimension formed by the clockwise (CW, ↑) and counterclockwise (CCW, ↓) modes at the same frequency. The coupling along the frequency dimension is achieved with a modulator (*13*). The coupling along the pseudospin dimension is achieved with a coupler in the shape of a figure eight (hereafter, "8-shaped coupler"), consisting of two-directional couplers connected by two nonintersecting waveguides. Our construction is different from methods of probing higher-dimensional phenomena using topological pumps, for which the physics with two extra dimensions has been explored in recent experiments (*11*, *12*). In these systems, a mathematical mapping between higher-dimensional lattices and lower-dimensional systems is achieved by varying some external parameters of the lower-dimensional system (*2*, *14*). Although signatures of higher-dimensional physics can be observed in such topological pumping schemes, the full dynamics are not captured because the external parameters are not the dynamical variables of the particles (*3*). In contrast, our approach provides the ability to explore physical dynamics in higher-dimensional space.

[1]Ginzton Laboratory and Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA. [2]Department of Applied Physics, Stanford University, Stanford, CA 94305, USA. [3]State Key Laboratory of Advanced Optical Communication Systems and Networks, School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, China. [4]Key Laboratory of Artificial Micro- and Nano-structures of Ministry of Education and School of Physics and Technology, Wuhan University, Wuhan 430072, China.
*Corresponding author. Email: yuanluqi@sjtu.edu.cn (L.Y.); shanhui@stanford.edu (S.F.)

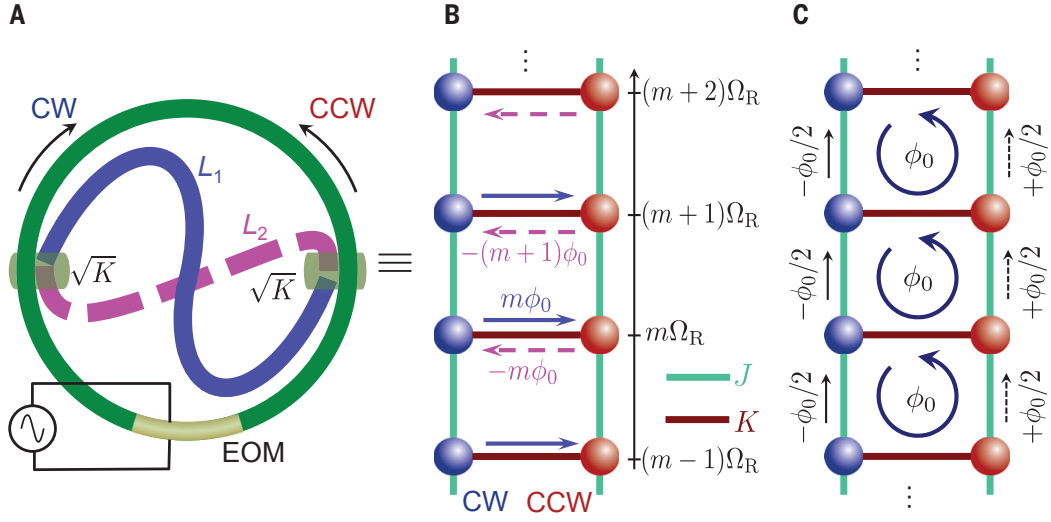


Fig. 1. A modulated ring resonator with CW-CCW mode-coupling and its corresponding lattice in synthetic dimensions. (**A**) Schematic of the ring of length L_0 with electro-optic modulation (EOM) and CW-CCW coupling. The CW and CCW modes form the pseudospin degree of freedom. The longitudinal modes of the ring separated by the FSR Ω_R form the frequency degree of freedom. The two directional couplers are connected into an 8-shaped coupler by two connecting waveguides of unequal lengths L_1 and L_2. By varying $\Delta L = L_1 - L_2$, the phases of couplings between CW and CCW modes [in (B) and (C)] can be varied, and hence a controllable effective magnetic field penetrates the ladder. The corresponding synthetic lattice is shown in two equivalent gauges: (**B**) a gauge with real inter-rung coupling J but complex inter-leg coupling (Eq. 1) and (**C**) a translationally invariant gauge (Eq. 2) with real inter-leg coupling K and complex inter-rung coupling.

The tight-binding Hamiltonian describing our system is

$$H = -\sum_{m,s}[\omega_m a^\dagger_{m,s}a_{m,s} + \sum_m J_{mm'}(t)a^\dagger_{m,s}a_{m',s}] - \sum_m K a^\dagger_{m,\uparrow}a_{m,\downarrow}e^{im\phi_0} + \text{H.c.} \quad (1)$$

where $a_{m,s}$ and $a^\dagger_{m,s}$ are the annihilation and creation operators for the mth longitudinal cavity mode with frequency $\omega_m = m\Omega_R$ and with pseudospin $s \in \{\downarrow, \uparrow\}$, and H.c. is the Hermitian conjugate. $J_{mm'}(t)$ is the coupling along the synthetic frequency dimension (*6*, *13*, *15*, *16*), produced by the electro-optic modulation (*7*). Because a small portion of the ring is modulated, this coupling can be simplified as $J_{mm'}(t) = J\cos\Omega_R t$; that is, the mode m can couple to all other modes of the system, and the coupling strength is independent of the mode indices (*7*). Here, Ω_R is the free spectral range (FSR), corresponding to the separation between the longitudinal modes. K in Eq. 1 is the strength of the coupling between the two legs of the ladder, created by the 8-shaped coupler comprising two-directional couplers with splitting amplitude $\sqrt{K}$. This coupling has a frequency- and direction-dependent phase $\pm m\phi_0$ (Fig. 1B), with $\phi_0 \sim \pi\Delta L/L_0$ (*17*), where ΔL is the length difference between the two connecting waveguides, and L_0 is the length of the ring. To explain how this phase $\pm m\phi_0$ is introduced, we note that the connecting waveguide depicted by the blue solid line in Fig. 1A couples exclusively from the CW to the CCW mode, whereas the connecting waveguide depicted by the dashed line couples only from the CCW to the CW mode. The phase difference between the coupling in the two directions is therefore $\Delta\phi(\omega) = \phi_{\downarrow\to\uparrow} - \phi_{\uparrow\to\downarrow} = \beta(\omega)\,\Delta L$, where $\beta(\omega)$ is the propagation constant at frequency ω for a mode in the connecting waveguides. Assuming that the connecting waveguides are the same as the waveguide of the ring, and because $\beta(\omega_m) = 2\pi m/L_0$, the phase difference $\Delta\phi$ increases linearly with m: $\Delta\phi(\omega_m) = 2\pi m\Delta L/L_0 = 2m\phi_0$.

To transform Eq. 1 into a time-independent Hamiltonian, we define $b_{m,\uparrow} = a_{m,\uparrow}e^{-im(\Omega_R t+\phi_0/2)}$ and $b_{m,\downarrow} = a_{m,\downarrow}e^{-im(\Omega_R t-\phi_0/2)}$ and use the rotating-wave approximation to get

$$H = -\frac{J}{2}\sum_m \left(b^\dagger_{m+1,\downarrow}b_{m,\downarrow}e^{i\phi_0/2} + b^\dagger_{m+1,\uparrow}b_{m,\uparrow}e^{-i\phi_0/2}\right) - K\sum_m b^\dagger_{m,\uparrow}b_{m,\downarrow} + \text{H.c.} \quad (2)$$

This Hamiltonian describes a two-legged ladder pierced by a uniform magnetic field (a Hall ladder) (*18*), as each plaquette is threaded by an effective magnetic flux ϕ_0 (Fig. 1, B and C). Thus, by choosing a nonzero ΔL, our structure in Fig. 1A naturally implements an effective magnetic field. Large magnetic fluxes spanning the entire range in $[-\pi, \pi]$ are achievable

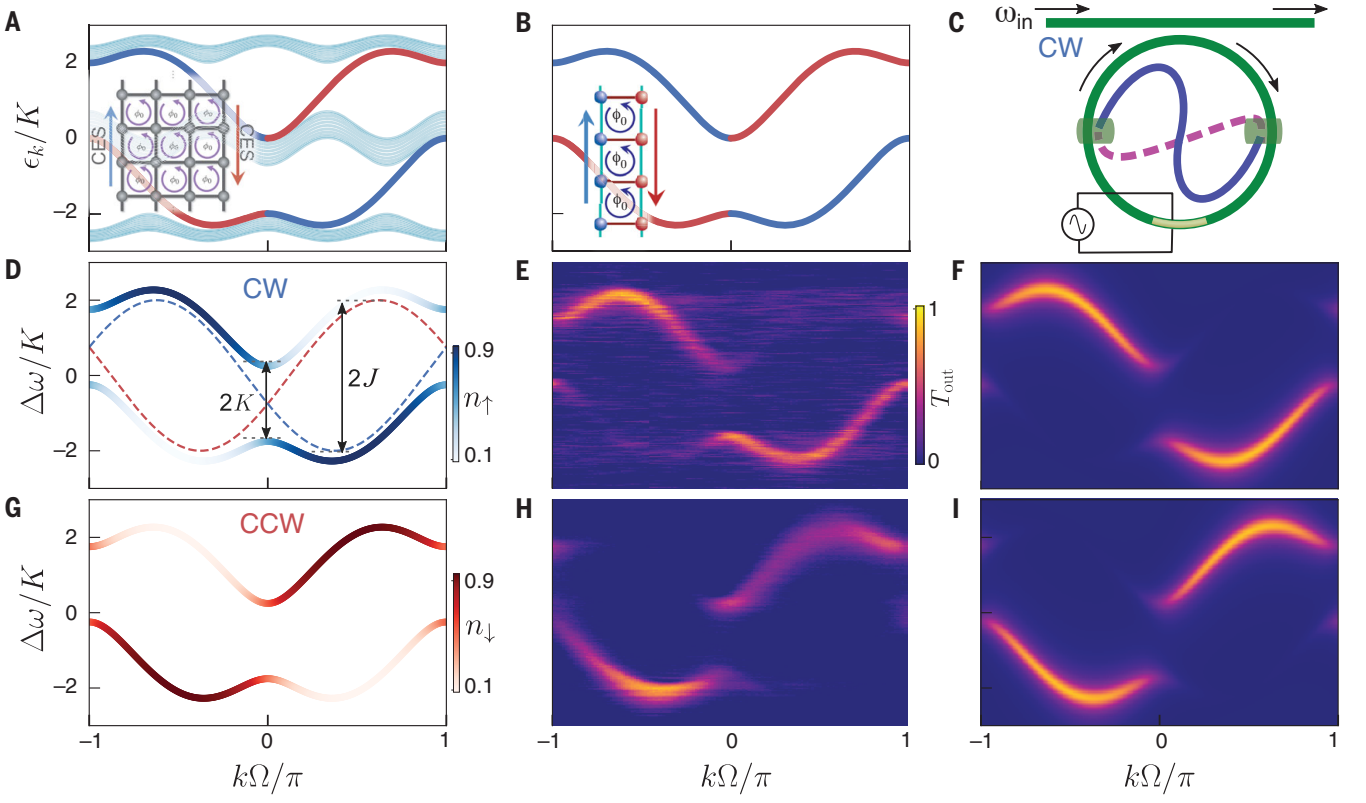


Fig. 2. Chiral band structure and spin-momentum locking in the synthetic Hall ladder. (**A**) Projected band structure of a 2D quantum Hall insulator infinite along the vertical direction and finite along the horizontal direction (as shown in the inset), showing topological chiral edge states (CES) highlighted in blue and red between the bulk band gaps. $\phi_0 = 2\pi/3$. (**B**) Band structure of the two-legged synthetic Hall ladder from the tight-binding Hamiltonian $\mathcal{H}(k)$ (Eqs. 2 and 3) for $J/K = 2$. The bulk bands disappear but signatures of chiral edge states are preserved (*19*). (**C**) Schematic setup to directly measure band structure by coupling an input-output waveguide to the ring in Fig. 1A. By varying ω_{in} and detecting the time-resolved transmission through the ring, the band structure can be directly read out in experiments. The CW (CCW) spin-resolved band structure can be detected by exciting the waveguide from the left (right) and recording its transmission. (**D** and **G**) Theoretical band structures, with color-coded pseudospin projections $n_\uparrow$ and $n_\downarrow$ for corresponding eigenstates. For the lower band, $k > 0$ states have predominantly CW pseudospin character, signifying spin-momentum locking. The dashed lines are band structures for the same J but for $K = 0$. (**E** and **H**) Experimental time-resolved transmission through the ring for CW excitation (E) and CCW excitation (H). $\Delta\omega$ is the detuning of the input frequency ω_{in} from the resonance frequency of the uncoupled CW and CCW modes. (**F** and **I**) Theoretical time-resolved transmission based on Floquet analysis (*17*). Experimental parameters: $J/2\pi = 1.95$ MHz and $K/2\pi = 0.97$ MHz. $\phi_0 \approx 3\pi/4$. Cavity linewidth $\gamma/2\pi = 480$kHz.

by choosing appropriate $\Delta L/L_0$. Because a purely one-dimensional (1D) lattice does not permit magnetic field effects, our system corresponds to the simplest lattice model where the physics emerging from effective magnetic fields for photons can be observed.

Instead of describing the system in Fig. 1 as a two-legged ladder threaded by a uniform magnetic field, we can alternatively derive the physics of this system in terms of magnetic field–controlled spin-orbit coupling (SOC), with the CW and CCW modes of each ring representing up and down spins. Going to the quasimomentum space (k-space), the Hamiltonian in Eq. 2 becomes $H = \int dk \mathbf{b}_k^\dagger \mathcal{H}(k) \mathbf{b}_k$, with $\mathbf{b}_k = \sqrt{\Omega/2\pi} \sum_m e^{im\Omega k} (b_{m,\uparrow}, b_{m,\downarrow})^T$ (where T denotes the transpose), and

$$\mathcal{H}(k) = -J\left[\mathbf{1}_2 \cos k\Omega \cos\frac{\phi_0}{2} + \sigma_z \sin k\Omega \sin\frac{\phi_0}{2}\right] - K\sigma_x \qquad (3)$$

Here σ_x, σ_y, and σ_z are Pauli matrices. To make the SOC explicit, we recast Eq. 3 into the form $\mathcal{H}(k) = \epsilon(k) \cdot \mathbf{1} + \mathbf{B}_{\mathrm{SOC}}(k) \cdot \boldsymbol{\sigma}$, where $\epsilon(k) = J\cos k\Omega \cos(\phi_0/2)$, $\mathbf{B}_{\mathrm{SOC}} = [K, 0, J \sin k\Omega \sin(\phi_0/2)]$, and $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$. The z component of $\mathbf{B}_{\mathrm{SOC}}$ depends on the quasimomentum k, signifying SOC. The degree of SOC is controlled by the effective magnetic flux ϕ_0. With the control of the magnetic flux, therefore, our system can exhibit a rich set of physics. Here we discuss three experimental observations of such physics, all controlled by the magnetic gauge potential: spin-momentum locking in the band structure, chiral currents, and a Meissner-to-vortex phase transition.

The Hall ladder has been formally shown to exactly reproduce the energies and eigenstates of the topological chiral edge modes of a 2D quantum Hall insulator (Fig. 2A) described by the Hofstadter model (*19*). Even if the entire bulk lattice sites are removed, the strip of plaquettes forming the ladder retains the chiral currents and spin-momentum locking, as can be seen by comparing Fig. 2B to Fig. 2A. This attests to the topological robustness of the 2D quantum Hall insulator. Such signatures of topological chiral edge modes are evident in the theoretically calculated band structure of $\mathcal{H}(k)$ (Fig. 2, D and G) along with the corresponding color-coded pseudospin projections $n_\uparrow = \cos^2(\theta_B/2)$ and $n_\downarrow = \sin^2(\theta_B/2)$, respectively. Here $\theta_B = \arctan\{K/[J \sin k\Omega \sin(\phi_0/2)]\}$ represents the chiral Bloch angle of the eigenstate, and its k-dependence signifies chiral spin-momentum locking (*19*, *20*): In the lower band, positive (or negative) k states have predominantly CW (or CCW) pseudospin character.

To directly detect the chiral modes of the Hall ladder, we use time-resolved band structure spectroscopy (*7*). We can selectively excite the CW or CCW pseudospin by exciting the waveguide from the left or right, respectively, and measure the transmitted signal to map out the band structure projected onto the corresponding spin (*17*). The results of these measurements (Fig. 2, E and H) were obtained by using a setup consisting of a fiber ring with an embedded electro-optic modulator and an 8-shaped coupler. The modulator is driven at $\Omega = 2\Omega_R = 29.6$ MHz [see (*17*) and (*21*) for details on the setup]. The measured band structure agrees with that from the tight-binding model (Fig. 2, D and G) and also with simulations using a rigorous Floquet analysis (Fig. 2, F and I) (*17*). This constitutes a measurement of the dispersion of chiral one-way states in synthetic dimensions. It is analogous to direct methods of exploring surface-state dispersions in SOC topological insulators [using angle-resolved photoemission spectroscopy (ARPES)] (*22*) or analyzing helical edge state dispersions in real-space photonic crystals (*23*). Spin-momentum locking is clearly seen in the experimental data (Fig. 2C), as the CW mode transmission predominantly peaks at positive quasimomenta for the lower band. Additionally, we observe that the direction of spin-momentum locking switches for the upper band.

The Hall ladder exhibits chiral currents—in our system, the CW (CCW) pseudospin evolves preferentially to higher- (lower-) frequency modes for the lower band. The direction of the current switches for the upper band. To quantify the direction of such spin- and band-dependent frequency evolution, we define the steady-state chiral current as

$$j_C = \sum_{m>m_L} P(m,\uparrow) - \sum_{m<m_L} P(m,\uparrow) \qquad (4)$$

where m_L is the order of the ring resonance closest to the input laser ($|\omega_{in} - m_L\Omega| < \Omega_R/2$) and $P(m,\uparrow)$ is the steady-state photon number of the CW mode at frequency $m\Omega$. To measure j_C, we use frequency- and spin-resolved heterodyne detection of the modal photon numbers in the lattice (*17*). Specifically, we frequency-shift a portion of the input laser by $\delta\omega$ = 500 MHz using an acousto-optic modulator and interfere it with the cavity output. Here $\delta\omega \gg |m - m_L|\Omega$ for all of the modes that we consider. A fast Fourier transform (FFT) of this interferogram directly yields $P(m)$. Heterodyne detection (i.e., the use of a frequency shift, as mentioned above) is essential. If one were to set $\delta\omega = 0$ in the experiment described

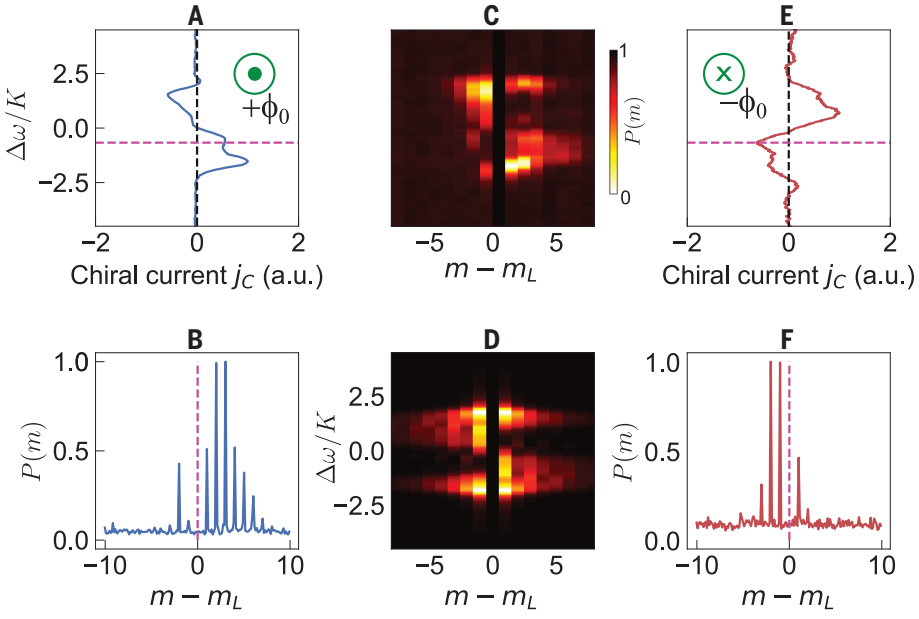


Fig. 3. Direct measurements of chiral currents in the synthetic Hall ladder through heterodyne detection. (**A**) Chiral current j_C (Eq. 4) versus laser-cavity detuning $\Delta\omega$ measured by heterodyne mixing the cavity output field with a frequency-shifted part of the input laser. The full heterodyne signal is shown in (C). The lower band shows a positive j_C for the CW mode. a.u., arbitrary units. (**B**) Steady-state–normalized photon number of the modes at frequencies $m\Omega$ in the lower band, at $\Delta\omega/K = -0.67$ indicated by the purple dashed line in (A). The asymmetric frequency mode occupation verifies that the CW mode predominantly evolves toward higher frequencies in the lower band. (**C**) Experimental heterodyne spectra mapping out the steady-state photon numbers for all $\Delta\omega$. (**D**) Theoretically calculated photon numbers based on a Floquet analysis. (**E** and **F**) Same as in (A) and (B), but with the direction of the effective magnetic field flipped, which causes a change in the sign of j_C. (A) and (C) also reveal a switching of the direction of chiral current on moving from the lower to the upper band. In (C) and (D), the strong signal in the excited mode ($m - m_L = 0$) has been suppressed to reveal the occupation of other modes clearly.

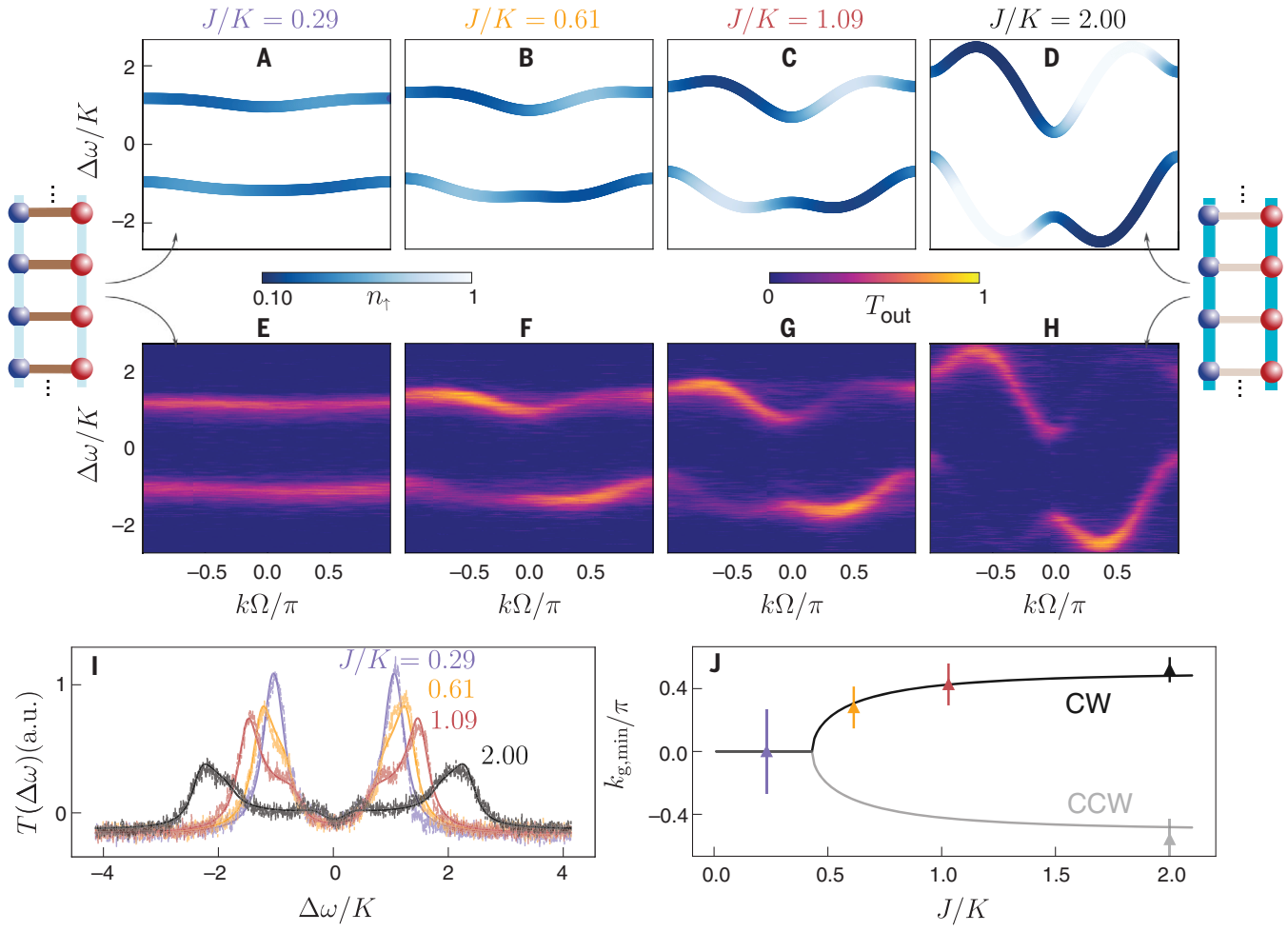


Fig. 4. Observation of phase transition through spin-resolved band structure measurements. (**A** to **D**) Theoretical band structure for $\phi_0 = 2.38 \approx 3\pi/4$, for increasing J/K. (**E** to **H**) Corresponding experimentally measured time-resolved transmission, showing good agreement with theory. The ladder insets on the left and right are indicative of the strengths in the pseudospin and frequency axes. J can be continuously tuned by varying the amplitude of the modulation signal. (**I**) Time-averaged transmission revealing the DOS. Van Hove singularities due to a diverging DOS are also visible in the transmission, smeared out by the cavity decay rate $\gamma/K = 0.37$. (**J**) Bifurcation of the energy minimum in k. Data points represent experimentally estimated splittings for band structures shown in (E) to (H), which agree with the solid lines based on Eq. 5.

above, one could not distinguish between the photon numbers at $m_L + m$ and $m_L - m$ modes, because they produce beat notes at the same radio frequency $m\Omega$. Figure 3A shows the measured chiral current j_C versus the laser detuning $\Delta\omega$. For each $\Delta\omega$, j_C is calculated from the heterodyne FFT spectrum. An example of such a spectrum at $\Delta\omega/K = -0.67$ is shown in Fig. 3B. In Fig. 3C, we show such spectra for all $\Delta\omega$. In Fig. 3D, we show a theoretical computation of the same spectrum. The overall shape of the theoretical spectrum agrees with the experiments. In both the theory and experimental results, the higher-frequency modes have a larger occupation ($j_C > 0$) in the lower band. The sign of j_C is switched for the upper band. Alternately, the sign of j_C can be switched by changing the direction of the effective magnetic field (Fig. 3, E and F), which corresponds to exchanging the lengths L_1 and L_2 in our system in Fig. 1A. Whereas our experiments measured the steady-state chiral current, we present theoretical simulations of chiral one-way propagation and the reversal of its direction with a switching of the magnetic flux in movies S1 to S3 (*17*). Our simulations show that such one-way propagation is resilient to backscattering around corners in a finite synthetic lattice for nontrivial fluxes $\phi_0 \neq \pi, 0$ but undergoes strong backreflection for trivial fluxes $\phi_0 = \pi, 0$ (movies S4 and S5) (*17*).

The Hall ladder in ultracold atomic systems has been predicted to exhibit a phase transition on increasing J/K, from a phase that has a single energy minimum in the ground state ("Meissner" phase) at $k = 0$ to a state that has a pair of energy minima at degenerate k points ("vortex" phase) (*19*, *24*). We demonstrate a similar transition in the band structure to illustrate the freedom in our system for shaping photonic bands. We adopt the same terminology to facilitate the comparison with existing literature. In our system, J can be easily tuned by changing the modulation voltage while K remains constant. For $J/K \ll 1$, the system can be described as a set of decoupled rungs of the ladder. In this regime, the eigenstates are the standing-wave symmetric and antisymmetric supermodes, resulting in flat bands split by $2K$ (Fig. 4, A, E, and left inset). Both bands have equal contributions from the CW and CCW legs of the ladder. For $J/K \gg 1$, the two legs of the ladder become decoupled, and we approach the sinusoidal band structure of a 1D tight-binding model with nearest-neighbor coupling (*7*). In the intermediate regime, the competition between synthetic SOC and effective magnetic field causes a transition in the band structure

from a single minimum at $k = 0$ (Fig. 4, A and E) to two minima (Fig. 4, B to H) at (*19*)

$$k_{g,\min} = \pm\arcsin\sqrt{\sin^2\frac{\phi_0}{2} - \frac{K^2}{J^2\tan^2(\phi_0/2)}} \quad (5)$$

The experimentally estimated band minima positions agree with the theoretical prediction within measurement uncertainties (Fig. 4J).

By measuring the time-averaged transmission instead of the time-resolved transmission, we detect the spin-projected density of states (DOS) (Fig. 4I). For $J \ll K, \gamma$, two peaks with Lorentzian lineshapes are seen, broadened by the cavity photon decay rate γ (Fig. 4I, blue curve). On increasing J, each of these peaks broadens owing to the increasing width of the corresponding band structure (orange curve). Eventually, additional peaks are visible for $J > 2\gamma$ (red and black curves), due to van Hove singularities at the edges of both energy bands (*20*, *25*).

Although some aspects such as spin-momentum locking, chiral currents, and van Hove singularities have been previously observed in atomic systems (*4*, *20*, *24*, *25*–*29*), several features are specific to our photonic implementation. First, we are able to directly measure the dispersion of the chiral one-way modes in synthetic space, owing to the time-resolved band structure spectroscopy technique, as opposed to mapping of the density-of-states in cold atom experiments (*20*, *25*). Second, we have access to the entire band structure, including the upper band, which allows us to experimentally observe the chirality switching (Fig. 3) when going from the lower to the upper band in a Hall ladder. Finally, our system exhibits frequency conversion, which can have applications in spectral manipulation of light. For example, the change in the dispersion relations associated with the Meissner-to-vortex phase transition can be used for tunable frequency conversion and frequency comb generation with tailorable spectral envelopes, as shown in fig. S5 (*17*). All of these features are achieved in a simple photonic structure consisting of a single modulated ring, completely based on the synthetic dimension concept.

Several new possibilities and future applications may be enabled using concepts demonstrated in our experiments. One notable possibility is the implementation of interacting Hamiltonians in synthetic frequency dimensions through the introduction of optical nonlinearities (*29*, *30*), which would permit the study of fundamental many-body physics (*31*) and have applications in quantum simulation (*1*) and quantum information processing. When combined with ideas from quantum photonics, one can generate high-dimensional hyperentanglement in pseudospin and frequency axes by exciting the CW and CCW modes with entangled photon pairs (*9*). Moreover, extensions of our setup can be used to manipulate photonic degrees of freedom, not only limited to frequency conversion but also including topologically protected manipulation of orbital angular momentum (*9*, *32*) and transverse spatial modes (*5*). The advent of nanophotonic lithium niobate microring modulators with bandwidths exceeding the ring FSR shows promise for realizing synthetic frequency dimensions on a chip (*33*). We anticipate that similar synthetic space concepts could be extended to other frequency ranges, such as microwaves (*34*), or to real-space photonic systems in which SOC (*35*), chiral quantum emission, and spin-momentum locking have been reported (*36*, *37*). Additionally, CW-CCW mode excitation in microrings has been explored for the study of non-Hermitian physics (*38*), counterpropagating solitons (*39*, *40*), and topological insulator lasers (*41*). These ideas can be combined with our experimentally demonstrated concepts of gauge potentials, effective magnetic fields, and SOC to manipulate and control light in versatile ways.

REFERENCES AND NOTES

1. O. Boada, A. Celi, J. I. Latorre, M. Lewenstein, *Phys. Rev. Lett.* **108**, 133001 (2012).
2. L. Yuan, Q. Lin, M. Xiao, S. Fan, *Optica* **5**, 1396–1405 (2018).
3. T. Ozawa, H. M. Price, *Nat. Rev. Phys.* **1**, 349–357 (2019).
4. M. Mancini *et al*., *Science* **349**, 1510–1513 (2015).
5. E. Lustig *et al*., *Nature* **567**, 356–360 (2019).
6. B. A. Bell *et al*., *Optica* **4**, 1433–1436 (2017).
7. A. Dutt *et al*., *Nat. Commun.* **10**, 3122 (2019).
8. A. Regensburger *et al*., *Nature* **488**, 167–171 (2012).
9. L. Yuan *et al*., *Phys. Rev. Lett.* **122**, 083903 (2019).
10. I. Martin, G. Refael, B. Halperin, *Phys. Rev. X* **7**, 041008 (2017).
11. M. Lohse, C. Schweizer, H. M. Price, O. Zilberberg, I. Bloch, *Nature* **553**, 55–58 (2018).
12. O. Zilberberg *et al*., *Nature* **553**, 59–62 (2018).
13. L. Yuan, Y. Shi, S. Fan, *Opt. Lett.* **41**, 741–744 (2016).
14. M. Wimmer, H. M. Price, I. Carusotto, U. Peschel, *Nat. Phys.* **13**, 545–550 (2017).
15. T. Ozawa, H. M. Price, N. Goldman, O. Zilberberg, I. Carusotto, *Phys. Rev. A* **93**, 043827 (2016).
16. C. Qin *et al*., *Phys. Rev. Lett.* **120**, 133901 (2018).
17. See supplementary materials.
18. K. Fang, Z. Yu, S. Fan, *Nat. Photonics* **6**, 782–787 (2012).
19. D. Hügel, B. Paredes, *Phys. Rev. A* **89**, 023619 (2014).
20. S. Kolkowitz *et al*., *Nature* **542**, 66–70 (2017).
21. A. Dutt *et al*., *ACS Photonics* **6**, 162–169 (2019).
22. D. Hsieh *et al*., *Nature* **452**, 970–974 (2008).
23. N. Parappurath, F. Alpeggiani, L. Kuipers, E. Verhagen, arXiv:1811.10739 [physics.optics] (26 November 2018).
24. M. Atala *et al*., *Nat. Phys.* **10**, 588–593 (2014).
25. L. F. Livi *et al*., *Phys. Rev. Lett.* **117**, 220401 (2016).
26. L. W. Cheuk *et al*., *Phys. Rev. Lett.* **109**, 095302 (2012).
27. F. A. An, E. J. Meier, B. Gadway, *Sci. Adv.* **3**, e1602685 (2017).
28. H. Cai *et al*., *Phys. Rev. Lett.* **122**, 023601 (2019).
29. M. E. Tai *et al*., *Nature* **546**, 519–523 (2017).
30. L. Yuan, A. Dutt, M. Qin, S. Fan, X. Chen, arXiv:1909.12466 [physics.optics] (27 September 2019).
31. R. Ma *et al*., *Nature* **566**, 51–57 (2019).
32. X.-W. Luo *et al*., *Nat. Commun.* **6**, 7704 (2015).
33. C. Reimer, Y. Hu, A. Shams-Ansari, M. Zhang, M. Loncar, arXiv:1909.01303 [physics.optics] (3 September 2019).
34. C. W. Peterson, W. A. Benalcazar, M. Lin, T. L. Hughes, G. Bahl, *Phys. Rev. Lett.* **123**, 063901 (2019).
35. D. L. P. Vitullo *et al*., *Phys. Rev. Lett.* **118**, 083601 (2017).
36. S. Barik *et al*., *Science* **359**, 666–668 (2018).
37. K. Y. Bliokh, D. Smirnova, F. Nori, *Science* **348**, 1448–1451 (2015).
38. S. Malzard, H. Schomerus, *Phys. Rev. A* **98**, 033807 (2018).
39. Q.-F. Yang, X. Yi, K. Y. Yang, K. Vahala, *Nat. Photonics* **11**, 560–564 (2017).
40. C. Joshi *et al*., *Opt. Lett.* **43**, 547–550 (2018).
41. M. A. Bandres *et al*., *Science* **359**, eaar4005 (2018).

ACKNOWLEDGMENTS

We thank D.A.B. Miller for initial discussions on the experiment and for providing lab space and equipment. **Funding:** This work is supported by a Vannevar Bush Faculty Fellowship (grant N00014-17-1-3030) from the U.S. Department of Defense and by MURI grants from the U.S. Air Force Office of Scientific Research (grants FA9550-17-1-0002 and FA9550-18-1-0379). L.Y. acknowledges support from the National Natural Science Foundation of China (grant 11974245). M.M. acknowledges support from the Swiss National Science Foundation (grant P300P2_177721). **Author contributions:** Q.L. developed the idea, in conjunction with L.Y. and A.D. A.D. designed, built, and characterized the setup and collected the data, in consultation with L.Y., Q.L., and S.F. M.M. and Q.L. contributed to theoretical analysis, simulations, and interpretations, with input from S.F., M.X., and L.Y. All authors contributed to discussion of the results and manuscript writing. S.F. supervised the project. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6473/59/suppl/DC1
Materials and Methods
Figs. S1 to S5
References (*42*–*44*)
Movies S1 to S5

29 August 2019; accepted 13 November 2019
Published online 28 November 2019
10.1126/science.aaz3071

TOPOLOGICAL MATTER

Absence of evidence for chiral Majorana modes in quantum anomalous Hall-superconductor devices

Morteza Kayyalha[1]*, Di Xiao[1]*, Ruoxi Zhang[1]*, Jaeho Shin[1], Jue Jiang[1], Fei Wang[1], Yi-Fan Zhao[1], Run Xiao[1], Ling Zhang[1], Kajetan M. Fijalkowski[2,3], Pankaj Mandal[2,3], Martin Winnerlein[2,3], Charles Gould[2,3], Qi Li[1], Laurens W. Molenkamp[2,3], Moses H. W. Chan[1]†, Nitin Samarth[1]†, Cui-Zu Chang[1]†

A quantum anomalous Hall (QAH) insulator coupled to an s-wave superconductor is predicted to harbor chiral Majorana modes. A recent experiment interprets the half-quantized two-terminal conductance plateau as evidence for these modes in a millimeter-size QAH-niobium hybrid device. However, non-Majorana mechanisms can also generate similar signatures, especially in disordered samples. Here, we studied similar hybrid devices with a well-controlled and transparent interface between the superconductor and the QAH insulator. When the devices are in the QAH state with well-aligned magnetization, the two-terminal conductance is always half-quantized. Our experiment provides a comprehensive understanding of the superconducting proximity effect observed in QAH-superconductor hybrid devices and shows that the half-quantized conductance plateau is unlikely to be induced by chiral Majorana fermions in samples with a highly transparent interface.

Topological superconductors (TSCs) are predicted to host Majorana fermions, particles that are their own antiparticles (*1*–*5*). These Majorana fermions obey non-Abelian statistics and are promising candidates to form a topological qubit, which is the basis for fault-tolerant topological quantum computation (*6*–*8*). TSCs are predicted to appear in a variety of condensed matter quantum systems including strong spin-orbit–coupled semiconductor-SC hybrid devices (*9*, *10*), fractional quantum Hall (QH) systems at filling factor $\nu = 5/2$ (*11*, *12*), spinless $p_x + ip_y$ SCs such as Sr_2RuO_4 (*2*, *13*), hybrid topological insulator (TI)–SC devices (*9*), integer QH insulators covered by a conventional s-wave SC (*14*), and thin films of transition metal dichalcogenides (*15*, *16*). Theoretical work has predicted a chiral TSC phase when a quantum anomalous Hall (QAH) insulator, a zero–magnetic field manifestation of the integer QH effect (*17*, *18*), is coupled to an s-wave SC (*14*, *19*).

The QAH effect has been experimentally demonstrated in thin films of magnetically doped TI (*18*, *20*–*22*). He *et al*. (*23*) recently reported a half-quantized plateau in the two-terminal conductance $\sigma_{1,2}$ converted from resistance measured across a millimeter-size QAH-Nb hybrid structure and interpreted the half-quantized $\sigma_{1,2}$ plateau during magnetization reversal as a "distinct signature" of one-dimensional chiral Majorana edge modes (CMEMs) (*19*). Alternative interpretations, however, are also possible. For example, Huang *et al*. (*24*) and Ji and Wen (*25*) theoretically discussed two different scenarios in which a $\sigma_{1,2} = 0.5e^2/h$ plateau (where e is the elementary charge and h is the Planck constant) can arise without invoking the Majorana physics. Huang *et al*. considered the percolation of QAH edges induced by magnetic disorder in the QAH insulator as an alternative origin for the $\sigma_{1,2} = 0.5e^2/h$ plateau. Ji and Wen argued that the $\sigma_{1,2} = 0.5e^2/h$ plateau can arise if the SC layer provides good electrical contact to the chiral edge modes of the QAH insulator. In other words, the local equilibrium between the chiral edge modes of the QAH insulators and the SC strip ensures that the total resistance is the series resistance of two separate QAH regions, each with h/e^2 resistance (*26*).

Here, we studied the effect of contact transparency in the appearance of the $\sigma_{1,2} = 0.5e^2/h$ plateau. To this end, we fabricated magnetic TI-SC hybrid devices, an example of which is shown in Fig. 1, A and B. Our device consists of a superconducting Nb strip (width ~20 μm) covering the entire width of the QAH layer on the left, a configuration similar to that in (*23*), and a narrow Nb finger (width ~200 nm) on the right (Fig. 1, A and B). The QAH sample in this device is a sandwich of 3QL Cr-doped $(Bi, Sb)_2Te_3$ / 5QL $(Bi, Sb)_2Te_3$ / 3QL Cr-doped $(Bi, Sb)_2Te_3$, where QL stands for quintuple layer (*27*). Our device was designed such that (i) the contact transparency between the magnetic TI and SC layers can be characterized using a differential conductance measurement on the QAH-Nb finger junction (*28*); (ii) the possible existence of the CMEMs can be investigated by analyzing the two-terminal conductance $\sigma_{1,2}$ deduced from resistance measured across the QAH-Nb strip device (*19*, *23*). Furthermore, our QAH film (i.e., magnetic TI) can be tuned to the metallic state using the back-gate voltage V_g. This allows us to probe the Andreev reflection involved in the magnetic TI-SC hybrid device through the entire phase diagram—that is, as a function of the chemical potential (tuned by V_g) and the external magnetic field. When the QAH layer is tuned into the metallic phase, we observed a strong enhancement of the zero-bias electrical conductance, nearly twice (~180%) the normal-state conductance presumably induced by Andreev reflection. The observation of Andreev reflection in our junction is strong evidence for the induced superconducting pair potential in the magnetic TI layer and allows us to study the effect of a transparent interface on the two-terminal conductance $\sigma_{1,2}$ in the QAH-SC hybrid structure. When the magnetic TI is in the QAH regime, the differential conductance is dominated by the density of state modulation (i.e., breakdown) of the QAH effect. When the QAH and SC layers are strongly coupled, as demonstrated by our differential conductance data, $\sigma_{1,2}$ is always half-quantized when the magnetization is well aligned. Our conclusions are supported by measurements on ~30 devices (*29*).

Figure 1C shows the temperature dependence of the Nb finger and the Nb strip resistance. The Nb finger becomes superconducting below its critical temperature $T_{c,finger}$ ~ 5 K. The critical temperature of the Nb strip $T_{c,strip}$ is ~8.6 K. Because we are using a two-terminal technique to measure the resistance (between electrodes 8a and 8b in Fig. 1A), the normal leads contribute ~40 ohms to the total resistance, which has been subtracted. Figure 1D plots the magnetic field ($\mu_0 H$) dependence of the resistances of the Nb finger and the Nb strip. The Nb strip has an upper critical field $\mu_0 H_{c2,strip}$ ~ 2.9 T. Shown in Fig. 1, E and F, are the $\mu_0 H$ dependence of the longitudinal resistance (conductance) ρ_{xx} (σ_{xx}) and the Hall resistance (conductance) ρ_{yx} (σ_{xy}) at $V_g = V_g^0$ = +42 V and T = 30 mK, where typical QAH characteristics, quantized ρ_{yx} (σ_{xy}) accompanied by vanishing ρ_{xx} (σ_{xx}), are observed. Because the ρ_{xx} peak value during magnetization reversal is comparable to the quantized ρ_{yx} value, the zero–Hall conductance $\sigma_{xy} = 0$ plateau [i.e., Chern number $C = 0$ phase (*30*)] is not observable. The $\sigma_{xy} = 0$ plateau is usually observed in thinner, uniformly doped QAH samples with a larger ρ_{xx} peak (*31*, *32*).

We characterized the interface transparency of the magnetic TI-Nb finger junction by measuring its differential conductance, which is related to the probabilities of the Andreev reflection (AR) and the normal reflection (NR) across the interface. Figure 2, A and B, shows the differential upstream conductance $\sigma_U = dI_{6,8}/dV_{7,8}$ and the downstream conductance $\sigma_D = dI_{6,8}/dV_{9,8}$, where the subscript

[1]Department of Physics, Pennsylvania State University, University Park, PA 16802, USA. [2]Faculty for Physics and Astronomy (EP3), University of Würzburg, Am Hubland, D-97074 Würzburg, Germany. [3]Institute for Topological Insulators, Am Hubland, D-97074 Würzburg, Germany.
*These authors contributed equally to this work.
†Corresponding author. Email: mhc2@psu.edu (M.H.W.C.); nxs16@psu.edu (N.S.); cxc955@psu.edu (C.-Z.C.).

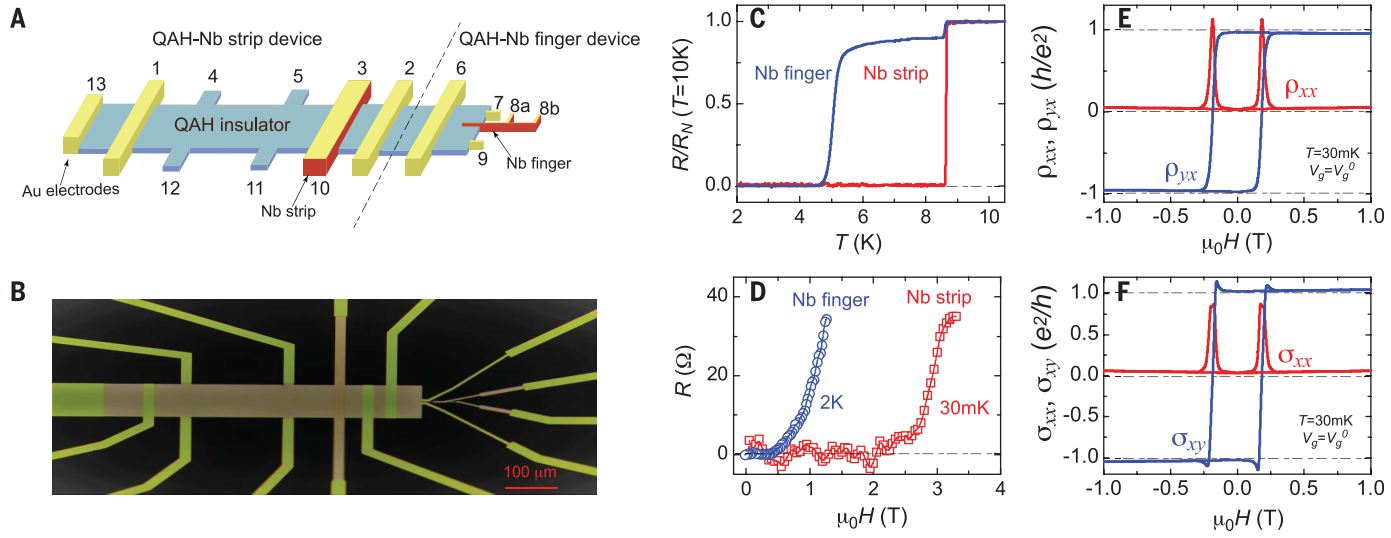


Fig. 1. QAH-Nb device and characterizations. (**A**) Schematic of the device consisting of a QAH insulator layer, a Nb strip, and a Nb finger. The Nb finger is used to characterize the magnetic TI-Nb contact transparency, whereas the Nb strip is used to study the two-terminal conductance $\sigma_{1,2}$ across the QAH-Nb structure. (**B**) Optical microscope image of the device. (**C**) Temperature dependence of the normalized resistance of the Nb finger and Nb strip. The drop in the resistance of the Nb finger at $T \sim 8.6$ K is associated with a superconducting transition of the Nb section with a larger width (~4 μm) in the device; see (B) and the inset of Fig. 2A. (**D**) $\mu_0 H$ dependence of the resistance of the Nb finger and Nb strip. (**E** and **F**) The four-terminal longitudinal and Hall resistance (ρ_{xx} and ρ_{yx}) (E) and their corresponding longitudinal and Hall conductance (σ_{xx} and σ_{xy}) (F) as a function of $\mu_0 H$ measured at $V_g = V_g^0 = +42$ V and $T = 30$ mK.

numbers correspond to the electrodes shown in Fig. 1A, at different magnetic fields. σ_U and σ_D are normalized by their respective values at $T > T_{c,finger}$ (i.e., σ_{6K}). For $V_g = V_g^0$, the differential conductance is determined by an interplay between AR and NR at the interface, as well as the breakdown of the QAH system (*33*–*35*); the breakdown of the QAH state turns out to be the dominant contribution in our samples (see fig. S3). On the other hand, σ_U (σ_D) is a better probe of the AR/NR ratio when the magnetic TI is in its metallic phase, as discussed below. To characterize the magnetic TI-Nb interface transparency, we applied a negative $V_g = -50$ V to reach the metallic phase of the magnetic TI. At zero magnetic field, we observed an enhancement of the zero-bias conductance approaching 180% of its high-temperature value, revealing a highly transparent magnetic TI-SC interface. Remarkably, although the superconductivity in the Nb finger is suppressed for $\mu_0 H > 0.5$ T (Fig. 1B and fig. S1E), the magnetic TI-Nb contact transparency is unaffected (Fig. 2, A and B). For $\mu_0 H$ larger than the coercive field ($\mu_0 H_c \sim 0.06$ T) of the magnetic TI layer at $T = 2$ K, zero-bias σ_U is slightly reduced and zero-bias σ_D is slightly increased. The reduction of σ_U and enhancement of σ_D are likely results of the magnetization reversal in the magnetic TI layer around the $\mu_0 H_c$ regime (see fig. S4).

In our experiment, it is difficult to extract accurately the voltage drop across the magnetic TI-Nb junction because a large portion of the voltage drop appears across the resistive part of the magnetic TI layer (*29*). Therefore, we plot the differential conductance as a function of the dc current, I_{dc}, rather than the dc voltage, V_{dc}. Furthermore, the enhancement in σ_U (σ_D) may be a result of the combined contribution of the AR process across the magnetic TI-Nb interface and the metal-to-superconductor transition induced by the critical current of the Nb finger. We also note that the slight reduction of σ_U (σ_D) close to zero bias (i.e., the small zero-bias conductance dip) observed in Fig. 2 is a result of the interplay between the AR and NR at the interface, as predicted by the Blonder, Tinkham,

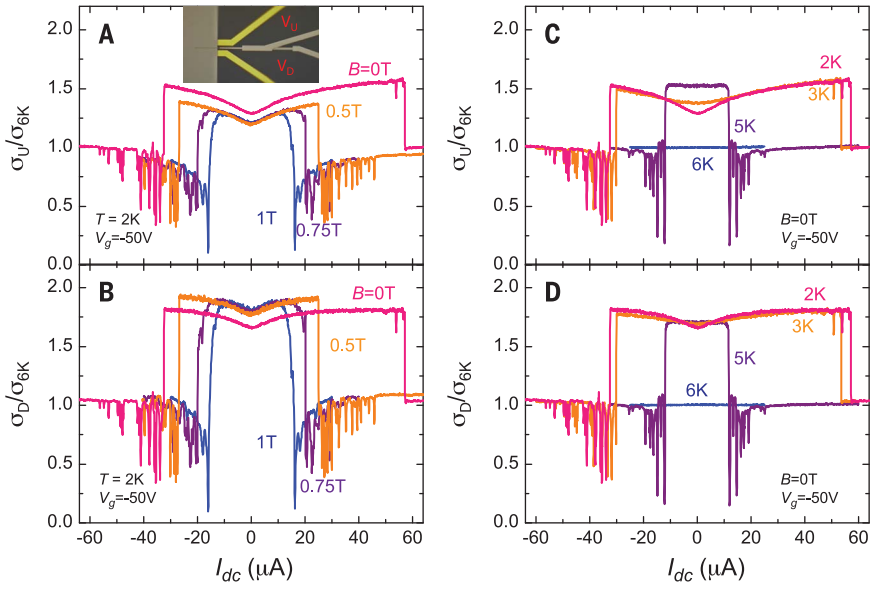


Fig. 2. Contact transparency in the magnetic TI-Nb finger device. (**A** and **B**) The differential upstream conductance $\sigma_U = dI_{6,8}/dV_{7,8}$ (A) and the downstream conductance $\sigma_D = dI_{6,8}/dV_{9,8}$ (B) of the magnetic TI-Nb finger junction normalized by their high-temperature ($T > T_{c,finger}$) values σ_{6K}, measured at different values of $\mu_0 H$ and $T = 2$ K. Inset of (A) shows a magnified optical image of the magnetic TI-Nb finger device. (**C** and **D**) The normalized σ_U (C) and σ_D (D) measured at different temperatures and zero magnetic field. The excitation current I_{ac} is 10 nA.

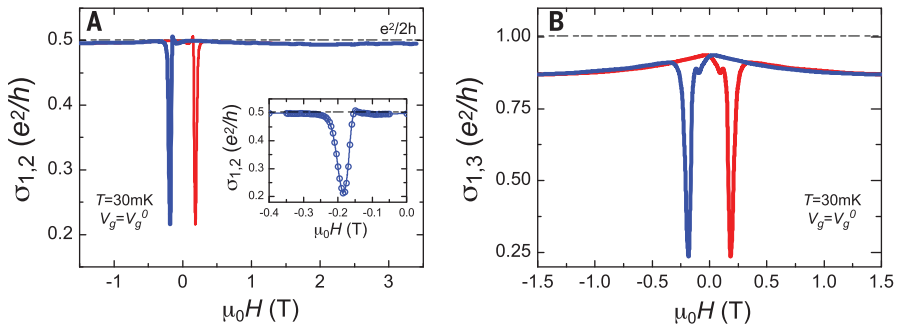


Fig. 3. Two-terminal conductance $\sigma_{1,2}$ across the QAH-Nb strip device. (**A**) μ_0H dependence of $\sigma_{1,2} = dI_{13,6}/dV_{1,2}$ measured at $V_g = V_g^0 = +42$ V and $T = 30$ mK. $\sigma_{1,2} \sim 0.5e^2/h$ for the entire μ_0H range when the magnetization is well aligned. No change in $\sigma_{1,2}$ is observed when the Nb strip transitions from the superconducting state to the normal state ($\mu_0H > \mu_0H_{c2,strip} \sim 2.6$ T). Inset magnifies the μ_0H axis during the magnetization reversal process. (**B**) μ_0H dependence of two-terminal conductance $\sigma_{1,3} = dI_{13,6}/dV_{1,3}$. $\sigma_{1,3}$ approaches $\sim e^2/h$ for $|\mu_0H| > \mu_0H_c$, indicating good contact transparency between the Nb strip and the QAH sample. The excitation current I_{ac} is 1 nA. Blue and red curves represent the process for decreasing and increasing μ_0H, respectively.

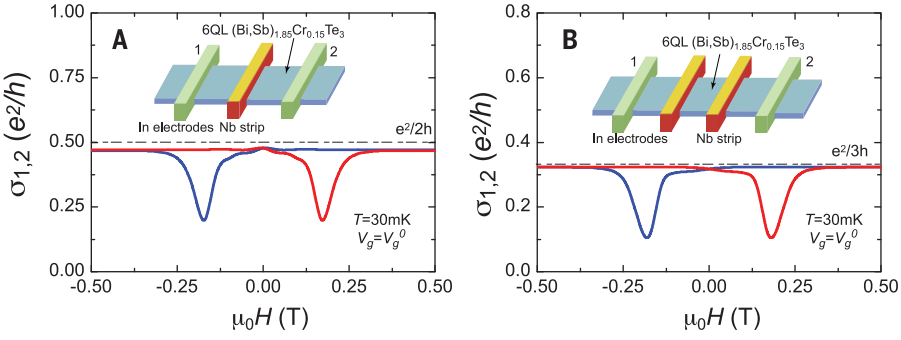


Fig. 4. Two-terminal conductance $\sigma_{1,2}$ in 6QL uniformly doped QAH-Nb strip devices. (**A**) μ_0H dependence of $\sigma_{1,2}$ measured across one Nb strip stacked on a 6QL (Bi, Sb)$_{1.85}$Cr$_{0.15}$Te$_3$ QAH sample at $V_g = V_g^0$ and $T = 30$ mK. (**B**) Same as (A) for two Nb strips. Insets show the corresponding device configurations. With increasing the number (n) of Nb strips, the corresponding $\sigma_{1,2}$ plateau decreases as $\sigma_{1,2} \sim e^2/(n+1)h$. The excitation current I_{ac} is 1 nA. Blue and red curves represent the process for decreasing and increasing μ_0H, respectively.

and Klapwijk (BTK) model for a normal metal-superconductor junction with a nonzero barrier height (*36*). To confirm that the sharp zero-bias conductance peak is indeed a result of the AR process at the interface, we studied the temperature dependence of σ_U (σ_D) versus I_{dc} in Fig. 2, C and D, where we observed a featureless σ_U (σ_D) at $T = 6$ K $> T_{c,finger}$. We note that at $T = 6$ K, the Nb finger is no longer superconducting (Fig. 1C), and thus the differential conductance is a sum of the contributions from the NR at the interface and the resistive part of the magnetic TI film. Therefore, the zero-bias conductance at $T = 6$ K takes the same value as that of the high-bias regime for $T \leq 5$ K, consistent with the AR picture for normal metal-superconductor junctions (*29*, *36*, *37*).

Our experimental observations reveal the presence of a highly transparent interface between the magnetic TI and Nb finger throughout the μ_0H range of interest (0 T $< \mu_0H <$ 1 T). Because the Nb finger and the Nb strip were deposited onto the QAH devices simultaneously, we expect the interface transparency across the magnetic TI-Nb junction to be similar for the strip and the finger. The transparent interface and the chiral nature of the edge modes in the QAH regime are expected to ensure that an electron propagating along the wide Nb strip will quickly become an equal mixture of electrons and holes (*28*).

Our QAH-SC hybrid device (minus the Nb finger) shown in Fig. 1, A and B, is similar to the device used in (*23*). A $\sigma_{1,2} \sim 0.5e^2/h$ plateau during the magnetization reversal ($\sim \mu_0H_c$) followed by a $\sigma_{1,2} \sim e^2/h$ plateau for the $\mu_0H_c < |\mu_0H| < \mu_0H_{c2,strip}$ regime is reported in (*23*). These measurements were interpreted as induced by the presence of the CMEMs; the transition from the $\sigma_{1,2} = e^2/h$ plateau to the $\sigma_{1,2} = 0.5e^2/h$ plateau was attributed to a topological phase transition in the TSC state from $N = 2$ to $N = 1$, where N denotes the number of CMEMs (*14*, *19*). In the same structure, an extremely small two-terminal conductance $\sigma_{1,3}$, measured between the Nb strip and the QAH sample, for $\mu_0H_c < |\mu_0H| < \mu_0H_{c2,strip}$ was also reported (*23*). The small value of $\sigma_{1,3}$ in this μ_0H range indicates that the Nb layer is likely decoupled from the QAH sample, and hence the $\sigma_{1,2} = e^2/h$ plateau may be a result of poor electrical contact between the QAH insulator and the Nb layers; in that case, there is no proximity-induced superconductivity and no AR at the QAH-Nb interface (*23*, *25*). We note that the observation of $\sigma_{1,2} = 0.5e^2/h$ in the QAH insulator phase is not unusual (*24*, *25*). Indeed, a normal metal (e.g., gold) overlaying the two edges of the QAH sample will give rise to such a quantization in $\sigma_{1,2}$ (*26*).

Our results from the QAH-SC strip devices can be explained without resorting to Majorana physics. Figure 3A displays the μ_0H dependence of the two-terminal conductance $\sigma_{1,2}$ for $V_g = V_g^0 = +42$ V of our device. In contrast to (*23*), we observed that $\sigma_{1,2} = dI_{13,6}/dV_{1,2} \sim 0.5e^2/h$ over the entire range of the magnetic field except in the μ_0H range, when the magnetization of the sample is being reversed near μ_0H_c. In this range, $\sigma_{1,2}$ drops to $\sim 0.2e^2/h$. Specifically, no change in $\sigma_{1,2}$ is observed when μ_0H is increased across the critical field of the Nb strip; that is, $\mu_0H_{c2,strip} \sim 2.9$ T (Fig. 1D). We also measured $\sigma_{1,3} = dI_{13,6}/dV_{1,3}$, the conductance between the Nb strip and the QAH sample (Fig. 3B). We found that $\sigma_{1,3} \sim e^2/h$ in the entire $|\mu_0H| > \mu_0H_c$ range, indicating that the Nb strip is strongly coupled to the QAH sample, leading to the equilibrium of chemical potentials between chiral edge modes of the QAH sample and bulk Nb layer (*25*). This behavior is what one would expect if a normal metal were used instead of the Nb strip. For $\mu_0H > \mu_0H_{c2,strip}$, the Nb strip turns into the normal state, hence $\sigma_{1,2}$ remains half-quantized. We have also studied 9QL V-doped TI samples, which were previously shown to exhibit perfect QAH effect (*35*, *38*–*40*) and signatures of axion electrodynamics (*38*). The devices were patterned using an optical lithography process and used MoRe as the SC strip. We again observed the $\sigma_{1,2} \sim 0.5e^2/h$ plateau for the entire μ_0H region with well-aligned magnetization (see fig. S8).

The existence of the zero–Hall conductance plateau with the $C = 0$ phase in a QAH sample was claimed as a prerequisite for the observation of the $N = 1$ TSC phase (*23*). The transition from the $C = 0$ (i.e., $N = 0$) phase to the $C = 1$ (i.e., $N = 2$) phase is given in (*23*) [citing (*19*)] as the key evidence for the existence of the $N = 1$ TSC phase. We note, however, that the theoretical calculations in

(*19*) treated the superconductor strip merely as the "source" of the small energy gap while overlooking the fact that the superconductor strip also serves as an "electrical short" for the QAH device.

Our results, on the other hand, show that the $\sigma_{1,2} = e^2/h$ plateau in the $C = 1$ phase is very likely a result of decoupling of the QAH insulator from the Nb layer. Hence, it is not predicated upon the existence of a TSC phase with $N = 2$. To exclude the possibility that the $\sigma_{1,2} \sim 0.5e^2/h$ plateau observed in our QAH sandwich sample may be caused by the absence of the zero–Hall conductance plateau (i.e., the $C = 0$ phase), we carried out measurements on QAH samples with the $C = 0$ phase. We fabricated two 6QL Cr-doped $(Bi, Sb)_2Te_3$ samples similar to the ones used in (*23*). Next, we scratched both samples into millimeter-size Hall bar structures (0.5 mm × 1 mm) and then sputtered Nb strips onto the samples with a mask. We measured $\sigma_{1,2}$ across one and two Nb strips. The value of $\sigma_{1,2}$ for one Nb strip sample (Fig. 4A) is similar to that measured in the QAH sandwich sample (Fig. 3A). Therefore, the existence or the nonexistence of the zero–Hall conductance plateau in QAH samples does not change our findings; specifically, the $\sigma_{1,2} \sim 0.5e^2/h$ plateau is observed for the entire $\mu_0 H$ region with well-aligned magnetization. We also studied the V-doped TI/TI/Cr-doped TI QAH sandwich samples, in which a well-established $C = 0$ insulating phase (i.e., the axion insulator state) emerges (*41*, *42*). Here, we also observed the $\sigma_{1,2} \sim 0.5e^2/h$ plateau for the entire $\mu_0 H$ region with well-aligned magnetization (see fig. S7). To better understand the relation between $\sigma_{1,2}$ and the coupling of the SC layer to the chiral edge modes, we measured $\sigma_{1,2}$ across a 6QL Cr-doped $(Bi, Sb)_2Te_3$ QAH sample with two Nb strips. We observed $\sigma_{1,2} \sim e^2/3h$ for the entire well-aligned $\mu_0 H$ regimes (Fig. 4B). The value of the $\sigma_{1,2}$ plateau decreases with an increasing number of Nb strips (n), specifically $\sigma_{1,2} \sim e^2/(n + 1)h$, which indicates that the total two-terminal resistance $\rho_{1,2}$ is a series resistance of $(n + 1)$ QAH sections, each contributing h/e^2 (*26*).

Our results demonstrate that if the SC layer is strongly coupled to the QAH sample, the two-terminal conductance $\sigma_{1,2}$ is half-quantized throughout the magnetic field range where the magnetization is well aligned. The agreement among the data obtained from the various QAH samples with different geometries demonstrates the robustness, reproducibility, and generality of the presented phenomena. Therefore, we conclude that the observation of $\sigma_{1,2} \sim 0.5e^2/h$ plateau alone is not sufficient evidence for the existence of chiral Majorana edge modes and the $N = 1$ TSC phase in the millimeter-size QAH-SC hybrid structures.

REFERENCES AND NOTES

1. E. Majorana, *Nuovo Cim.* **14**, 171–184 (1937).
2. N. Read, D. Green, *Phys. Rev. B* **61**, 10267–10297 (2000).
3. C. W. J. Beenakker, *Annu. Rev. Condens. Matter Phys.* **4**, 113–136 (2013).
4. J. Alicea, *Rep. Prog. Phys.* **75**, 076501 (2012).
5. X. L. Qi, S. C. Zhang, *Rev. Mod. Phys.* **83**, 1057–1110 (2011).
6. A. Y. Kitaev, *Ann. Phys.* **303**, 2–30 (2003).
7. F. Wilczek, *Nat. Phys.* **5**, 614–618 (2009).
8. C. Nayak, S. H. Simon, A. Stern, M. Freedman, S. Das Sarma, *Rev. Mod. Phys.* **80**, 1083–1159 (2008).
9. L. Fu, C. L. Kane, *Phys. Rev. Lett.* **100**, 096407 (2008).
10. R. M. Lutchyn, J. D. Sau, S. Das Sarma, *Phys. Rev. Lett.* **105**, 077001 (2010).
11. G. Moore, N. Read, *Nucl. Phys. B* **360**, 362–396 (1991).
12. A. Stern, *Nature* **464**, 187–193 (2010).
13. A. P. Mackenzie, Y. Maeno, *Rev. Mod. Phys.* **75**, 657–712 (2003).
14. X. L. Qi, T. L. Hughes, S. C. Zhang, *Phys. Rev. B* **82**, 184516 (2010).
15. N. F. Q. Yuan, K. F. Mak, K. T. Law, *Phys. Rev. Lett.* **113**, 097001 (2014).
16. Y. T. Hsu, A. Vaezi, M. H. Fischer, E. A. Kim, *Nat. Commun.* **8**, 14985 (2017).
17. F. D. M. Haldane, *Phys. Rev. Lett.* **61**, 2015–2018 (1988).
18. C. Z. Chang *et al.*, *Science* **340**, 167–170 (2013).
19. J. Wang, Q. Zhou, B. Lian, S. C. Zhang, *Phys. Rev. B* **92**, 064520 (2015).
20. C. Z. Chang *et al.*, *Nat. Mater.* **14**, 473–477 (2015).
21. J. G. Checkelsky *et al.*, *Nat. Phys.* **10**, 731–736 (2014).
22. X. Kou *et al.*, *Phys. Rev. Lett.* **113**, 137201 (2014).
23. Q. L. He *et al.*, *Science* **357**, 294–299 (2017).
24. Y. Y. Huang, F. Setiawan, J. D. Sau, *Phys. Rev. B* **97**, 100501 (2018).
25. W. Ji, X. G. Wen, *Phys. Rev. Lett.* **120**, 107002 (2018).
26. C. Z. Chang *et al.*, *Phys. Rev. Lett.* **115**, 057206 (2015).
27. J. Jiang *et al.*, arXiv 1901.07611 [cond-mat.mes-hall] (28 January 2019).
28. G. H. Lee *et al.*, *Nat. Phys.* **13**, 693–698 (2017).
29. See supplementary materials.
30. J. Wang, B. Lian, S. C. Zhang, *Phys. Rev. B* **89**, 085106 (2014).
31. X. Kou *et al.*, *Nat. Commun.* **6**, 8474 (2015).
32. Y. Feng *et al.*, *Phys. Rev. Lett.* **115**, 126801 (2015).
33. M. Kawamura *et al.*, *Phys. Rev. Lett.* **119**, 016803 (2017).
34. E. J. Fox *et al.*, *Phys. Rev. B* **98**, 075145 (2018).
35. M. Götz *et al.*, *Appl. Phys. Lett.* **112**, 072102 (2018).
36. G. E. Blonder, M. Tinkham, T. M. Klapwijk, *Phys. Rev. B* **25**, 4515–4532 (1982).
37. M. Tinkham, *Introduction to Superconductivity* (McGraw-Hill, 1996).
38. S. Grauer *et al.*, *Phys. Rev. Lett.* **118**, 246801 (2017).
39. M. Winnerlein *et al.*, *Phy. Rev. Mater.* **1**, 011201 (2017).
40. S. Grauer *et al.*, *Phys. Rev. B* **92**, 201304 (2015).
41. D. Xiao *et al.*, *Phys. Rev. Lett.* **120**, 056801 (2018).
42. M. Mogi *et al.*, *Sci. Adv.* **3**, eaao1669 (2017).
43. C. Z. Chang, Data for "Absence of Evidence for Chiral Majorana Modes in Quantum Anomalous Hall-Superconductor Devices"; https://doi.org/10.7910/DVN/QWQEHT, Harvard Dataverse (2019).

ACKNOWLEDGMENTS

We thank C. X. Liu, K. T. Law, B. Lian, J. Wang, X. Dai, J. Jain, H. Z. Lu, Z. Wang, B. H. Yan, G. H. Lee, Y. L. Chen, K. He, W. J. Ji, Q. K. Xue, and X. D. Xu for helpful discussions. **Funding:** Supported by ONR grant N-000141512370 and Penn State 2DCC-MIP under NSF grant DMR-1539916 (N.S.); DOE grant DE-FG01-08ER46531 (Q.L.); NSF grant DMR-1707340 (M.H.W.C.); NSF-CAREER award DMR-1847811, ARO Young Investigator Program Award W911NF1810198, and an Alfred P. Sloan Research Fellowship (C.-Z.C.); and EU ERC-AG Programs (project 3-TOP and 4TOPS) (C.G. and L.W.M.). Support for transport measurements and data analysis at Penn State is provided by DOE grant DE-SC0019064. **Author contributions:** N.S., M.H.W.C., and C.-Z.C. conceived and designed the experiment; D.X. and R.X. grew the QAH samples. M.K. fabricated the devices; M.K., J.S., and R.Z. performed the dilution refrigerator measurements; M.K., J.J., F.W., Y.-F.Z., and L.Z. carried out the PPMS transport measurements; C.G. and L.W.M. conceived and designed the experiment done in Würzburg; M.W. grew the QAH samples in Würzburg; K.M.F. and P.M. fabricated the devices and performed the dilution refrigerator measurements in Würzburg; M.K., C.G., N.S., M.H.W.C., and C.-Z.C. wrote the manuscript; all authors contributed to the analysis of the data and the final editing of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data in the main text and the supplementary materials are available at (*43*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6473/64/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S10
References (*44*–*46*)

9 April 2019; accepted 7 November 2019
10.1126/science.aax6361

SUPERCONDUCTIVITY

Atomic manipulation of the gap in $Bi_2Sr_2CaCu_2O_{8+x}$

F. Massee[1]*, Y. K. Huang[2], M. Aprili[1]

Single-atom manipulation within doped correlated electron systems could help disentangle the influence of dopants, structural defects, and crystallographic characteristics on local electronic states. Unfortunately, the high diffusion barrier in these materials prevents conventional manipulation techniques. Here, we demonstrate the possibility to reversibly manipulate select sites in the optimally doped high-temperature superconductor $Bi_2Sr_2CaCu_2O_{8+x}$ using the local electric field of the tip of a scanning tunneling microscope. We show that upon shifting individual Bi atoms at the surface, the spectral gap associated with superconductivity is seen to reversibly change by as much as 15 milli–electron volts (on average ~5% of the total gap size). Our toy model, which captures all observed characteristics, suggests that the electric field induces lateral movement of local pairing potentials in the CuO_2 plane.

One of the challenges in the study of high-temperature superconductivity in the cuprates (*1*) is their intrinsic inhomogeneous nature. This is exemplified by the archetypal system $Bi_2Sr_2CaCu_2O_{8+x}$ (Bi2212), whose complicated crystal structure includes an incommensurate structural supermodulation (*2*) and interstitial oxygen dopant atoms. Particularly notable is the large variation in spectral gap size over nanometer-scale distances (*3*–*5*), which has been shown to reflect local variations in the superconducting transition temperature T_c (*6*, *7*) and has been correlated to both oxygen dopants (*8*, *9*) and structural inhomogeneity (*10*). These correlations, however, typically involve averages over a large number of sites. To directly probe the influence of dopants, structural defects, and crystallographic characteristics on the local electronic states at the atomic scale, direct non-invasive control of the dopant positions and the inhomogeneous crystal structure they inhabit is therefore desirable. Unfortunately, common techniques for single-atom manipulation involving short-range forces between the tip and the atom (*11*–*13*) and/or vibrational excitation using the tunneling current (*14*–*16*) are not well suited for controllable manipulation of atoms in single-crystal cuprate materials. This is because the dopant atoms are buried under the surface and the diffusion barrier of the surface atoms themselves is too high. Alternatively, the electric field can be used to manipulate a surface (*17*–*19*); however, because the field profile depends on the size of the tip apex, this process is, in practice, difficult to control unless specific atoms are more sensitive to the field than others owing to their charge or local environment. We discovered that this is exactly the case in Bi2212, where we find two local environments that are more readily influenced by the electric field than the rest of the system.

Figure 1A shows a schematic of the two environments that allow for field-induced atom manipulation: surface Bi atoms on the crest of the periodic modulation of the bulk crystal structure (*2*), henceforth referred to as the supermodulation, and the recently discovered weakly coupled oxygen dopants (*20*). To manipulate the atoms, we position our tip above the surface and slowly increase the sample bias voltage, V_s. Upon reaching ~800 mV at a tunnel current of ~100 pA, we start to observe jumps in the current, where each jump corresponds to the manipulation of an atom below the tip. For manipulation voltages $V_s \geq 1.2$ V, the current typically shows two types of jumps (Fig. 1B): small ones corresponding to the manipulation of a near-surface oxygen dopant, which we can detect through their signature in the differential conductance (*8*, *9*, *20*), and big jumps corresponding to the manipulation of a surface Bi atom on the crest of the supermodulation that are observed directly in the constant current images (Fig. 1C) [section 1 of (*21*)]. We can freeze-in any new surface-and-dopant configuration by switching back to low bias voltage. Because the spatial extent of the highest electric field is determined by the size of the tip apex, which is typically a few tens of nanometers in diameter, in principle, several hundred Bi and oxygen dopants can be affected. However, unlike in a homogeneous system where the threshold field for manipulation is identical for all field-affected atoms or molecules (*19*), the existence of intrinsic lattice distortions, the supermodulation, and numerous local dopant environments in Bi2212 (*20*) leads to a range of threshold fields, enabling selective and reversible manipulation by carefully tuning the tip position and the manipulation voltage. As expected for electric field–induced manipulations, all manipulations we observe are confined to a roughly circular area with a diameter of a few tens of nanometers around the location of the tip during the application of the manipulation voltage (fig. S5). Owing to the aforementioned intrinsic inhomogeneous nature of Bi2212, studying the influence of single atoms on the local electronic structure is normally impossible, and averaging over a large number of sites is required instead. With the ability to locally rearrange a select number of atoms, however, we can effectively remove the inhomogeneous background information by considering the difference in local electronic

[1]Laboratoire de Physique des Solides, CNRS UMR 8502, Bâtiment 510, Université Paris-Sud, Université Paris-Saclay, 91405 Orsay, France. [2]Institute of Physics, University of Amsterdam, 1098XH Amsterdam, Netherlands.
*Corresponding author. Email: freek.massee@u-psud.fr

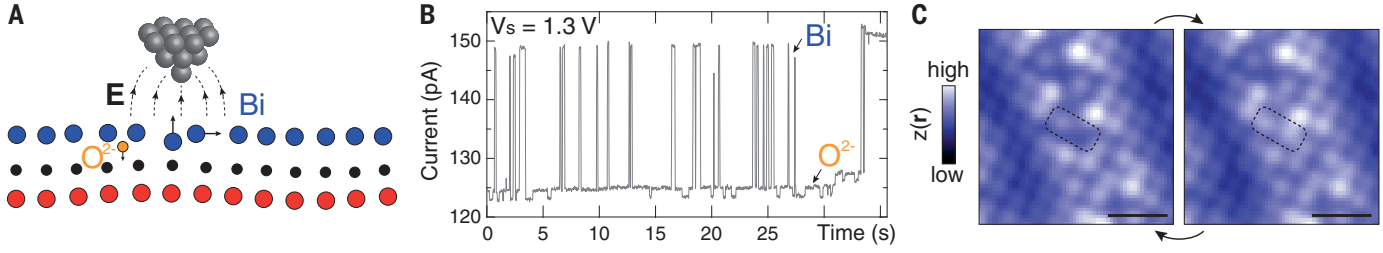


Fig. 1. Electric field–induced atom manipulation. (**A**) Illustration of the field-induced manipulation of surface Bi atoms and near-surface oxygen dopants (Bi, blue; O, black or orange; Cu, red). (**B**) Current as a function of time at V_s = 1.3 V. Small jumps signal oxygen dopant manipulation (O^{2-}), large jumps signal surface Bi rearrangement (Bi). (**C**) Constant current images, z(**r**), (E_{set} = −100 meV; I_{set} = 100 pA) of reversible surface Bi atom manipulation upon treatment with V_s > 1.2 V. The atoms in the dashed box are manipulated. Scale bars, 1 nm. All measurements throughout this work were performed at temperature $T \leq 4.2$ K.

states before and after manipulation. These difference measurements will only show contrast where the density of states has been altered by the atom manipulation, providing a direct link between single atomic sites and the local electronic structure.

To generate a difference map of the spectral gap, we record low-energy differential conductance maps and extract a map of the peak-to-peak gap, $\Delta_i(\mathbf{r})$, using identical settings before and after atom manipulation. Figure 2, A and B, shows one example of a gap map taken before and after ~10 dopants and Bi surface atoms have been manipulated. Three subsequent electric field manipulation treatments are shown in section 4 of (*21*). Whereas the topographies and the gap maps are mostly identical, excluding any change of the tip itself, select locations show substantial changes, with gaps increasing and decreasing by >10 meV. The two spectra in Fig. 2C, which are taken at the same location before and after the electric field–induced manipulation, highlight the significance of these changes. The spectra are predominantly affected at the peak energies, whereas the low energy states, $|E| < 20$ meV, are hardly modified, if at all, highlighting the insensitivity of the latter to disorder (*22*) (fig. S7). From the two gap maps, we then calculate the difference map, $D_{2-1}(\mathbf{r}) = \Delta_2(\mathbf{r}) - \Delta_1(\mathbf{r})$, which shows directly where the gap has enhanced (blue) or decreased (red) upon the electric field manipulation (Fig. 2D). As the histograms of the gap changes in Fig. 2E show, the average gap size over the entire field of view is preserved: The gap is both reduced and enhanced in equal measure in the vicinity of the manipulated atoms.

Using the topographies taken simultaneously with the gap maps, we can next pinpoint where surface Bi atoms have been manipulated (fig. S8 and S9); the locations of manipulated Bi atoms have been marked with black dots on the two gap maps (Fig. 2, A and B) and in their difference map in Fig. 2D. Similarly, from high energy differential conductance maps we can extract the position of all near-surface oxygen dopants, as well as determine which dopants have changed [section 1 of (*21*)]. The gap changes in Fig. 2D are clearly linked to the atomic manipulations, but in a rather unexpected manner: The altered Bi sites mark the boundary between regions of increasing and decreasing gaps, whereas the gap on the sites itself is hardly affected. Whereas one would expect the manipulation of oxygen dopants to have a strong effect on the gap, the contribution to the gap modifications of the near-surface dopants we manipulate does not appear to be the dominant one: The correlation between their location and where the gap changes is only moderate, and when we manipulate a single near-surface dopant atom, the gaps in its vicinity shift by a few milli–electron volts at most (fig. S6). This observation is in line with a previous study that found these dopants (i.e., those with a resonance at -1 eV $< E < -0.4$ eV) to have a much weaker correlation to the intrinsic gap inhomogeneity than the ones closer to the CuO_2 plane (*9*). The dominant contribution of manipulated Bi atoms and the secondary nature of the manipulated oxygen atoms are exemplified by the simulation shown in Fig. 2F. As discussed below, the simulation, which is in good agreement with experiment, uses a model that takes into account only Bi manipulations.

Atomic scale difference images of two consecutive manipulations of the same Bi atom are shown in Fig. 3A. Two observations stand out. First, whenever a surface modification reverts in a subsequent electric field treatment, the gap reverts back as well. Consequently, the difference images between the first and third map, $D_{3-1}(\mathbf{r})$, is featureless (fig. S11), and $D_{3-2}(\mathbf{r}) = -D_{2-1}(\mathbf{r})$ (Fig. 3A, right and left, respectively). For isolated manipulation events, even the full spectrum reverts to the original configuration (fig. S12). Second, the direction along which the gap is altered appears to be linked to the direction of the surface Bi atom repositioning. As can be seen from Fig. 3A and as shown schematically in Fig. 3B, each time a Bi atom moves down below (up into) the BiO plane, a neighboring atom laterally shifts toward (away) from it, leading to an enhancement of the spectral gap in the direction of this shift.

To quantify the direction of the gap change in more detail, we average the $D(\mathbf{r})$ images around each surface Bi manipulation after aligning their direction of maximum gap increase [see section 4 of (*21*)]. As Fig. 4A shows, a clear dipole profile, centered at the surface modification in an otherwise unaffected environment, is obtained. The two lobes of the dipole have opposite signs, are a few

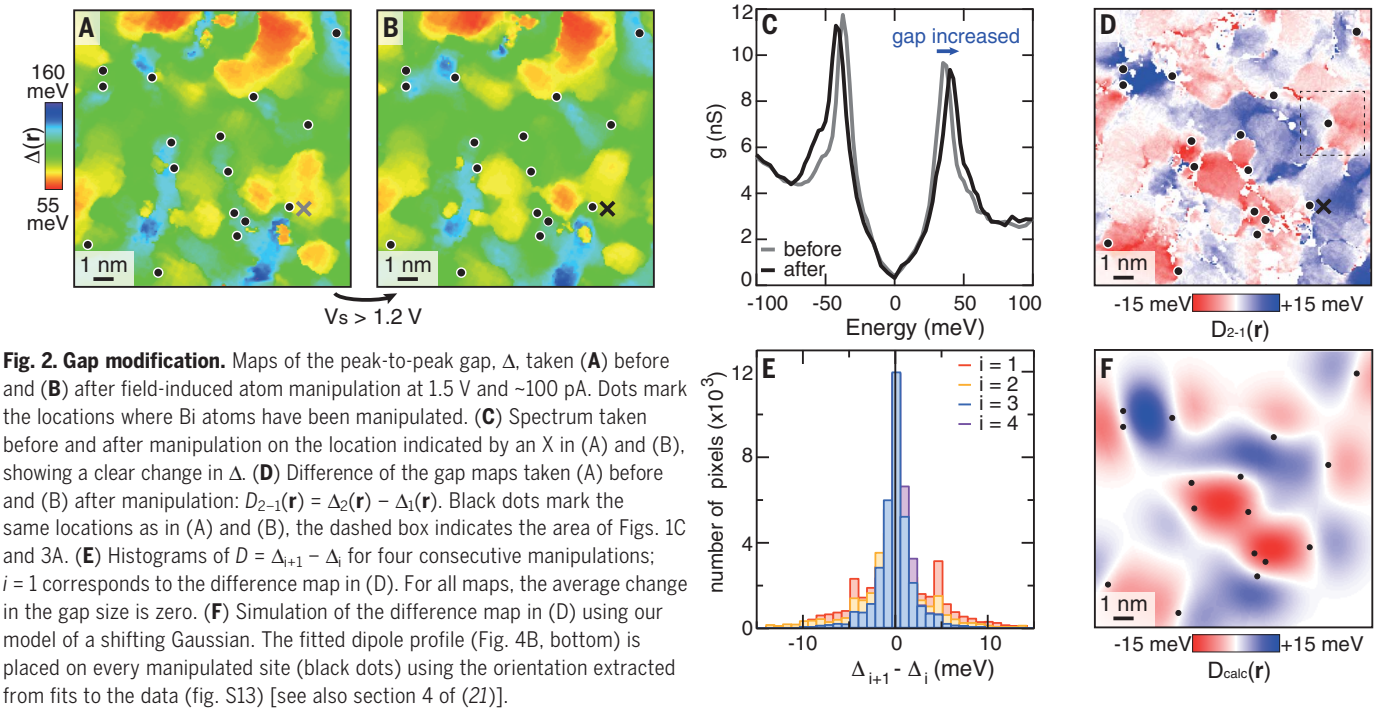


Fig. 2. Gap modification. Maps of the peak-to-peak gap, Δ, taken (**A**) before and (**B**) after field-induced atom manipulation at 1.5 V and ~100 pA. Dots mark the locations where Bi atoms have been manipulated. (**C**) Spectrum taken before and after manipulation on the location indicated by an X in (A) and (B), showing a clear change in Δ. (**D**) Difference of the gap maps taken (A) before and (B) after manipulation: $D_{2-1}(\mathbf{r}) = \Delta_2(\mathbf{r}) - \Delta_1(\mathbf{r})$. Black dots mark the same locations as in (A) and (B), the dashed box indicates the area of Figs. 1C and 3A. (**E**) Histograms of $D = \Delta_{i+1} - \Delta_i$ for four consecutive manipulations; $i = 1$ corresponds to the difference map in (D). For all maps, the average change in the gap size is zero. (**F**) Simulation of the difference map in (D) using our model of a shifting Gaussian. The fitted dipole profile (Fig. 4B, bottom) is placed on every manipulated site (black dots) using the orientation extracted from fits to the data (fig. S13) [see also section 4 of (*21*)].

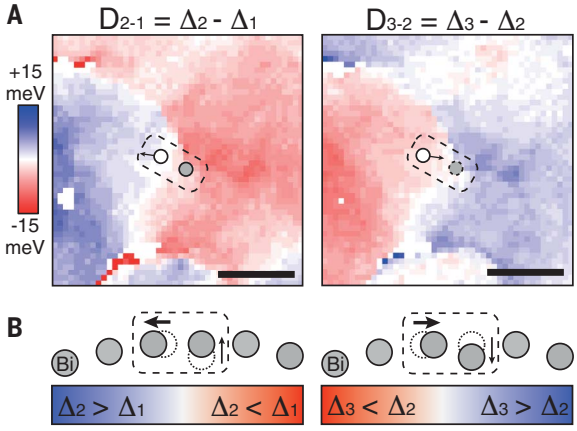


Fig. 3. Atomic position dependence of gap changes. (**A**) $D_{2-1} = \Delta_2(\mathbf{r}) - \Delta_1(\mathbf{r})$ image on the same area as the constant current images in Fig. 1C. The manipulated Bi atoms are marked by circles, the lateral movement upon manipulation is indicated by an arrow. The right panel shows the opposite gap modification upon a reversion of the atomic configuration after a subsequent V_s = 1.5 V manipulation. Scale bars, 1 nm. (**B**) Schematic of the surface change and its effect on the gap: One atom moves up (down), another shifts laterally away from (toward) it, resulting in gap enhancement in the direction of the shift. More details and examples can be found in sections 1 and 4 of (*21*).

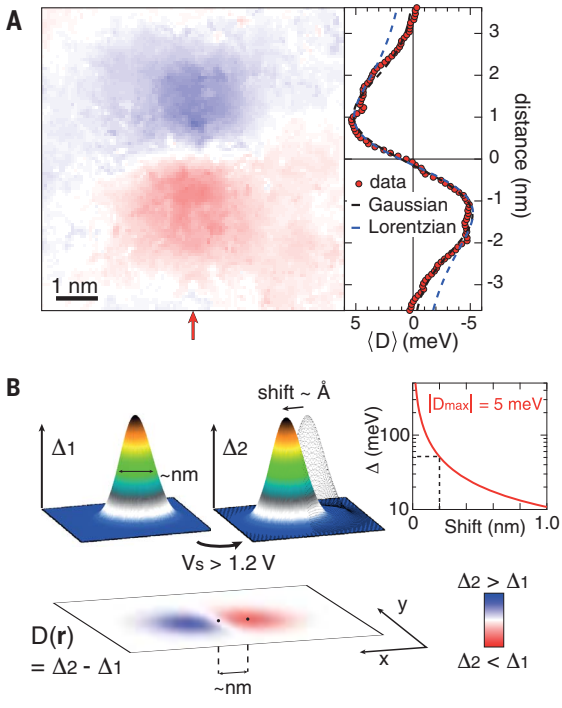


Fig. 4. Dipole profile and toy model. (**A**) Average $D(\mathbf{r})$ for all topographic modifications after aligning their orientation. The arrow indicates where the vertical line trace on the right is taken for the experimental data (red circles) and for the toy model using a Gaussian (black dashed line) and Lorentzian (blue dashed line) peak profile. (**B**) A 2D Gaussian (width = 1.7 nm; top left) is shifted by a fraction of a nanometer (top right, shift exaggerated for clarity), leading to a difference (bottom) upon their subtraction with the same shape and length scales as the experimental data in (A). The inset shows how the amplitude of the subtracted Gaussians depends on the shift distance: a 2-Å shift requires an amplitude of Δ ~ 50 meV (dashed lines).

nanometers in size, and are located at opposite ends, ~1.5 nm distance from the surface manipulation. The horizontal line-cut on the right side of Fig. 4A shows this in more detail. The most straightforward way to create a dipole-shaped profile in a difference image is by subtracting two identical [two-dimensional (2D)] peaks that are laterally shifted with respect to each other (Fig. 4B). In the context of our experiment, this would correspond to a single peak that laterally shifted by the electric field manipulation, which upon generating the difference map is thus subtracted from itself at a slightly different location. The maxima of the subtraction, i.e., the lobes of the dipole, are in the direction of the shift, as we observe in the experiment. For shifts smaller than the width of the peak, the lateral extent of the resulting dipole will depend exclusively on the width, whereas the magnitude of the dipole is a function of the size of the shift and the peak amplitude (Fig. 4B, inset). Using this simple toy model, we can accurately reproduce both the size and shape of the observed dipole when we use a Gaussian profile for the peak, as well as simulate the experimental data (Fig. 2F). Conversely, a Lorentzian profile decays too slowly to properly fit the tails of the difference image, as can be seen from the line-cut comparison in Fig. 4A. For a realistic lateral shift of 2 Å (see, for example, Fig. 1C), the Gaussian requires an amplitude of ~50 meV, which is not unreasonable given the average peak-to-peak gap size of Δ = 96 meV. Lastly, the width of the Gaussian is 1.7 nm, which is comparable to the superconducting coherence length.

A Gaussian-shaped gap distribution that falls off with length scales on the order of the coherence length is highly suggestive of a local pairing potential originating from a point-like object. When we laterally move this entity using the electric field of the tip, its local pairing potential moves with it, leading to a dipole-shaped feature in the difference image. The physical origin of the object is possibly related to the Bi atoms themselves, although the contribution of the Bi orbitals to the low energy density of states is limited (*23*). More likely, the apical oxygen atoms directly below the Bi atoms shift concomitant with the Bi manipulation, leading to the gap modifications. The importance of the apical oxygen atoms in both the tunneling process (*24*–*26*) and the gap size (*10*) have been stressed previously. Our observations provide additional input for further theoretical investigations—particularly those that take into account on-site correlations (*27*, *28*)—into the origin of the spectral gap, and the tunneling process, in this cuprate superconductor.

An alternative source of the dipole-shaped difference image could be the appearance of topological defects that introduce a 2π-rotation of the phase of the order parameter (*29*). However, creation and annihilation of topological defects has to occur in pairs. Given our finite field of view, it may be that one-half of the pair is outside our measurement range, but given the large field-of-view topographic before-and-after comparison (fig. S5), and assuming that both pairs will have a signature in topography, this is unlikely. Furthermore, in our optimally doped system we do not find a significant correlation between the gap changes and topological defects in the smectic, or the d-form factor density wave, reported for underdoped samples (*30*, *31*). Extension of our work to lower doping concentrations, where the various charge- and spin-ordered states are more predominant, should give a more definite answer to this issue. Additionally, the influence of dopants and the surface structure on these (ordered) states themselves can be studied directly using the field-induced atom manipulation we introduce in this work.

The observation of a profound influence on the peak-to-peak gap in tunneling experiments of subnanometer shifts in atomic positions highlights the importance of the lattice on the local electronic properties of the cuprates. The spatial profile of the gap modification we observe is highly suggestive of the field-induced lateral movement of a local pairing potential in the CuO_2 plane originating from a point-like object. This work demonstrates an avenue to

noninvasively and reversibly probe the influence of the local lattice on the electronic states of cuprate high-temperature superconductors and related compounds.

REFERENCES AND NOTES

1. B. Keimer, S. A. Kivelson, M. R. Norman, S. Uchida, J. Zaanen, *Nature* **518**, 179–186 (2015).
2. Y. Gao, P. Lee, P. Coppens, M. A. Subramania, A. W. Sleight, *Science* **241**, 954–956 (1988).
3. T. Cren, D. Roditchev, W. Sacks, J. Klein, *EPL* **54**, 84–90 (2001).
4. C. Howald, P. Fournier, A. Kapitulnik, *Phys. Rev. B* **64**, 100504 (R) (2001).
5. S. H. Pan *et al*., *Nature* **413**, 282–285 (2001).
6. K. K. Gomes *et al*., *Nature* **447**, 569–572 (2007).
7. A. N. Pasupathy *et al*., *Science* **320**, 196–201 (2008).
8. K. McElroy *et al*., *Science* **309**, 1048–1052 (2005).
9. I. Zeljkovic *et al*., *Science* **337**, 320–323 (2012).
10. J. A. Slezak *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 3203–3208 (2008).
11. D. M. Eigler, E. Schweizer, *Nature* **344**, 524–526 (1990).
12. L. Bartels, G. Meyer, K.-H. Rieder, *Phys. Rev. Lett.* **79**, 697–700 (1997).
13. M. Ternes, C. P. Lutz, C. F. Hirjibehedin, F. J. Giessibl, A. J. Heinrich, *Science* **319**, 1066–1069 (2008).
14. D. M. Eigler, C. P. Lutz, W. E. Rudge, *Nature* **352**, 600–603 (1991).
15. T.-C. Shen *et al*., *Science* **268**, 1590–1592 (1995).
16. T. W. Fishlock, A. Oral, R. G. Egdell, J. B. Pethica, *Nature* **404**, 743–745 (2000).
17. L. J. Whitman, J. A. Stroscio, R. A. Dragoset, R. J. Celotta, *Science* **251**, 1206–1210 (1991).
18. M. A. Rezaei, B. C. Stipe, W. Ho, *J. Chem. Phys.* **110**, 4891–4896 (1999).
19. M. Alemani *et al*., *J. Am. Chem. Soc.* **128**, 14446–14447 (2006).
20. F. Massee, Y. K. Huang, M. S. Golden, M. Aprili, *Nat. Commun.* **10**, 544 (2019).
21. Materials and methods are available as supplementary materials.
22. M. C. Boyer *et al*., *Nat. Phys.* **3**, 802–806 (2007).
23. M. S. Hybertsen, L. F. Mattheiss, *Phys. Rev. Lett.* **60**, 1661–1664 (1988).
24. I. Martin, A. V. Balatsky, J. Zaanen, *Phys. Rev. Lett.* **88**, 097003 (2002).
25. Y. He, T. S. Nunner, P. J. Hirschfeld, H.-P. Cheng, *Phys. Rev. Lett.* **96**, 197002 (2006).
26. J. Nieminen, H. Lin, R. S. Markiewicz, A. Bansil, *Phys. Rev. Lett.* **102**, 037001 (2009).
27. A. Kreisel *et al*., *Phys. Rev. Lett.* **114**, 217002 (2015).
28. P. Choubey, A. Kreisel, T. Berlijn, B. M. Andersen, P. J. Hirschfeld, *Phys. Rev. B* **96**, 174523 (2017).
29. A. Mesaros *et al*., *Science* **333**, 426–430 (2011).
30. M. J. Lawler *et al*., *Nature* **466**, 347–351 (2010).
31. K. Fujita *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **111**, E3026–E3032 (2014).
32. F. Massee, Supporting data for "Atomic manipulation of the gap in $Bi_2Sr_2CaCu_2O_{8+x}$" by F. Massee *et al.*, Version 1, Harvard Dataverse (2019); https://doi.org/10.7910/DVN/UUSS6R.

ACKNOWLEDGMENTS

We thank M. Civelli, A. Mesaros, and P. Simon for fruitful discussions. **Funding:** F.M. acknowledges funding from H2020 Marie Skłodowska-Curie Actions (grant 659247) and the ANR (ANR-16-ACHN-0018-01). **Author contributions:** F.M. and M.A. conceived of the study and discussed and interpreted the results. F.M. performed and analyzed all measurements. Y.K.H. grew the samples. F.M. wrote the manuscript with M.A. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data files for the results presented here are available at (*32*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6473/68/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S13
References (*33*–*38*)

8 February 2019; accepted 7 November 2019
10.1126/science.aaw7964

APPLIED PHYSICS

Topological mechanics of knots and tangles

Vishal P. Patil[1], Joseph D. Sandt[2], Mathias Kolle[2], Jörn Dunkel[1]*

Knots play a fundamental role in the dynamics of biological and physical systems, from DNA to turbulent plasmas, as well as in climbing, weaving, sailing, and surgery. Despite having been studied for centuries, the subtle interplay between topology and mechanics in elastic knots remains poorly understood. Here, we combined optomechanical experiments with theory and simulations to analyze knotted fibers that change their color under mechanical deformations. Exploiting an analogy with long-range ferromagnetic spin systems, we identified simple topological counting rules to predict the relative mechanical stability of knots and tangles, in agreement with simulations and experiments for commonly used climbing and sailing bends. Our results highlight the importance of twist and writhe in unknotting processes, providing guidance for the control of systems with complex entanglements.

Knots are among the oldest, most enduring human technologies, as valuable to ancient builders (*1*) and mariners (*2*) as to modern engineers and surgeons (*3*). Thought to predate the wheel (*1*), knotted structures owe their extraordinary longevity and widespread usage to an inherent mechanical robustness that arises from the subtle interplay of topology, elasticity, and friction. Over the course of many centuries, sailors, weavers, climbers, and surgeons have acquired a wealth of knowledge about the benefits and drawbacks of various types of knots (*1*, *2*). Experience has taught us that certain knots are more stable than others, but we are still largely unable to predict the mechanical behavior of knots and tangles from basic topological observables (*4*), such as the number and relative ordering of crossings. Although recent experimental and theoretical research has revealed important insights into the competition between force transmission and friction in special classes of knots (*5*, *6*), hitches (*7*), and fabrics (*8*), there currently exists no comprehensive mathematical theory (*9*) linking the topological and mechanical properties of knotted elastic structures.

Physical knots and their topology first assumed a central role in science with the introduction of Kelvin's vortex-atom model in the 1860s (*10*). Since then, the fundamental importance of entangled structures has become firmly established in a diverse range of disciplines and contexts (*9*). In physics, for example, interactions between knotted defect lines are essential to understanding and controlling dynamics and mixing in classical and complex fluids (*11*–*14*), including liquid crystals (*15*), plasmas (*16*), and quantum fluids (*17*). Whereas the energetic costs associated with topological transformations are typically low in liquids and gases (*17*), they tend to become prohibitively large in entangled solids (*5*, *6*). This fact has profound consequences for the stability and function of natural and engineered structures, from the microscopic knots in DNA (*18*, *19*), proteins (*20*–*22*), and polymers (*23*) to knitted clothes (*8*) and macroscopic meshworks (*24*). Achieving a unified understanding of these various systems requires taking into account not only their topological but also their elastic properties. Because key concepts from topology and elasticity theory remain applicable over a wide range of scales, deciphering the topological principles (*5*, *6*, *9*) that determine the mechanical stability of knots promises insights into a broad spectrum of physically entangled structures. Therefore, our main goal is to identify generic topological counting rules that enable us to estimate which members of a given family of elastic knots are the most robust against untying. To this end, we combined elements from mathematical and physical knot theory (*9*, *25*) with optomechanical experiments and quantitative continuum modeling (Fig. 1).

We were interested in tying two lines together so that they form a stable longer rope, a task known as "tying a bend" among sailors (*2*). Mathematically, this configuration describes an oriented 2-tangle, defined as the union of two oriented open curves embedded in space (*4*). Although an elegant mathematical formalism exists to describe certain simple families of 2-tangles (*26*), little is known theoretically about even the most basic bend knots used in practice. We constructed a topological phase diagram that explains the relative stability of a selection of bends that are commonly used in the sailing and climbing communities. To validate the underlying topological model, we compared its predictions with simulations of an optomechanically verified continuum theory and with quantitative measurements using laboratory "prisoner's escape" experiments.

Our optomechanical experiments use recently developed color-changing photonic fibers (*27*) that allow for the imaging of strain in knots (Fig. 1, A and B). These fibers derive structural color from a multilayer cladding composed of alternating layers of transparent elastomers with distinct refractive indices wrapped around an elastic core. Their coloration varies with the thickness of layers in the periodic cladding, which changes upon elongation or bending. As is typical of macroscopic materials at room temperature, the persistence lengths of the fibers used in our experiments are several orders of magnitude larger than the diameters of the tightened knots (*28*), with empirical knowledge (*2*) indicating that essential knot properties are only weakly dependent on the elastic modulus. Theoretically, we describe knotted fibers using a damped Kirchhoff model (*5*, *28*, *29*) validated through comparison with photographs depicting the strain-induced color changes in mechanoresponsive photonic fibers (Fig. 1, A to C). Simulating the tightening process of a 1-tangle, corresponding to a single knotted fiber pulled at both ends (Fig. 1, A and B, and movie S1), reveals the relative strengths and localization of the bending and stretching strains (Fig. 1, D and E), which are not individually discernible in our experiments. Furthermore, the Kirchhoff model highlights why topological considerations (*4*) alone do not suffice to explain the mechanical behaviors (*2*) of real-world knots: Loosening or tightening a knot transforms any of its planar projections according to a sequence of three elementary topology-preserving Reidemeister moves, R_1, R_2, and R_3 (Fig. 1F). Despite being topologically equivalent, the move R_1 is energetically distinct as it involves substantial changes in strain, whereas moves R_2 and R_3 are energetically favored soft modes (Fig. 1F), implying that physical knots preferentially deform by R_2 and R_3. Thus, to link the physical properties of tangled fibers to their topology, one must merge concepts from classical mathematical knot theory (*4*) with elasticity theory (*5*, *6*, *9*, *30*).

Continuum simulations provide guidance for how one can complement bare topological knot diagrams (*4*) with coarse-grained mechanical information that is essential for explaining why certain knots are more stable than others (Fig. 2). In contrast to a 1-tangle, which is tightened by pulling diametrically at its two ends (Fig. 2A), each strand of a bend knot has one pulled and one free end (Fig. 2B). Therefore, the local fiber velocity directions in the center-of-mass frame of the bend knot define natural fiber orientations on the underlying 2-tangle (Fig. 2B), thus establishing mapping between bends and oriented 2-tangles. At each contact crossing, the fibers mutually generate a frictional self-torque with well-defined handedness, depending on the relative velocity and ordering of the two fiber strands (Fig. 2, B

[1]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [2]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
*Corresponding author. Email: dunkel@mit.edu

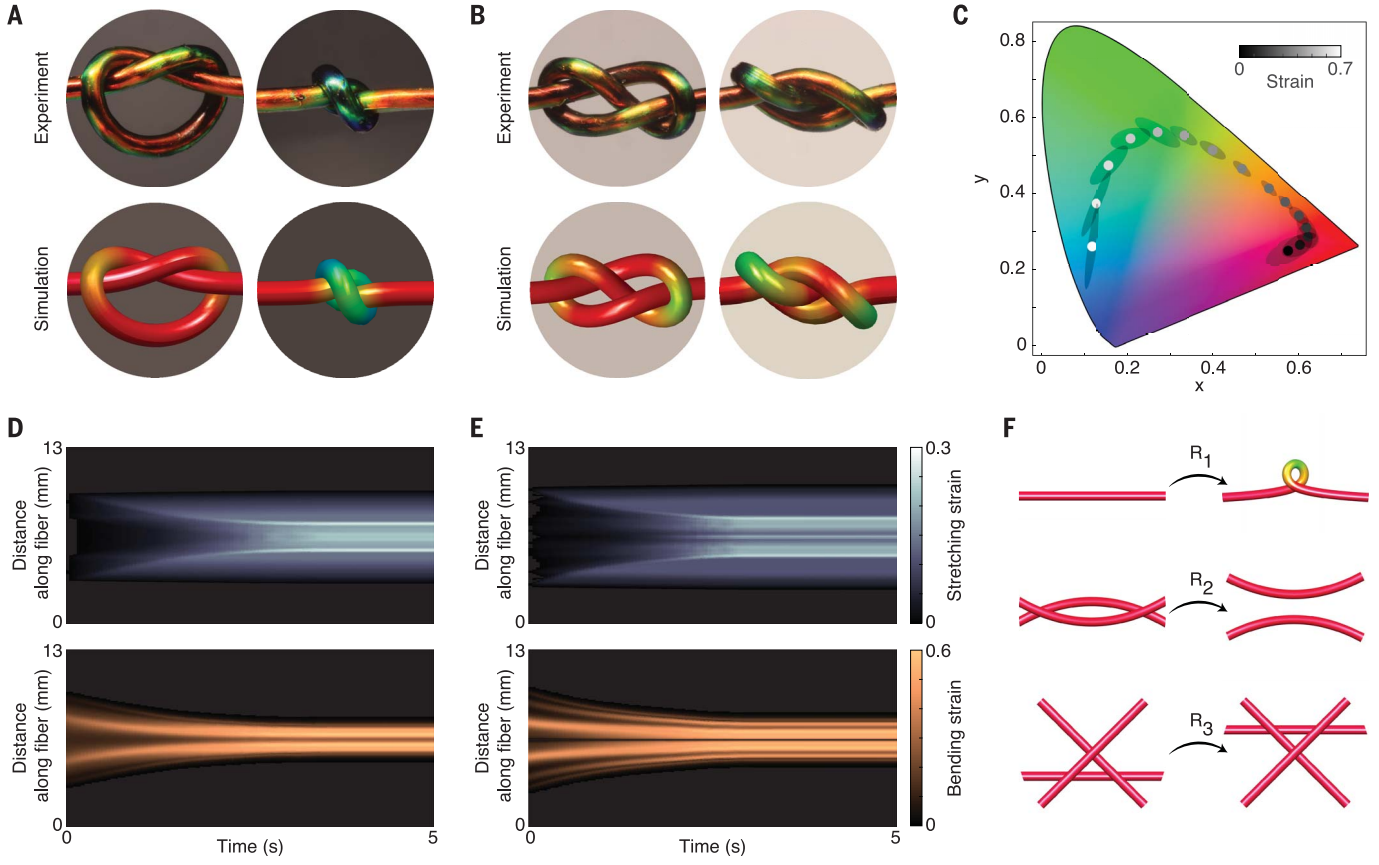


Fig. 1. Experiments and simulations reveal mechanical properties of knots. (**A** and **B**) Color-changing mechanoresponsive fibers confirm the stress patterns predicted by continuum simulations for the trefoil knot (A) and the figure-of-eight knot (B) during the tightening process (movie S1). Fiber diameter is 0.4 mm. (**C**) Dependence of fiber color on strain visualized as a trajectory in the CIE 1931 XYZ color space, where mean positions (solid circles) lie within standard deviation ellipses (*28*). This strain color coding is used in panels (A), (B), and (F). (**D** and **E**) Simulations revealing the relative strength of bending and stretching strains along knots. (D) and (E) show the evolution of these two complementary strain contributions during tightening of the trefoil knot in (A) and the figure-of-eight knot in (B). Pulling force is 0.02 N. The elastic moduli are given in (*28*). (**F**) Topology-preserving Reidemeister moves affect the elastic energy of the underlying fibers differently. Move R_1 induces strain and thus requires higher energy than R_2 and R_3, highlighting that both topological and elastic properties determine the mechanical behavior of knots.

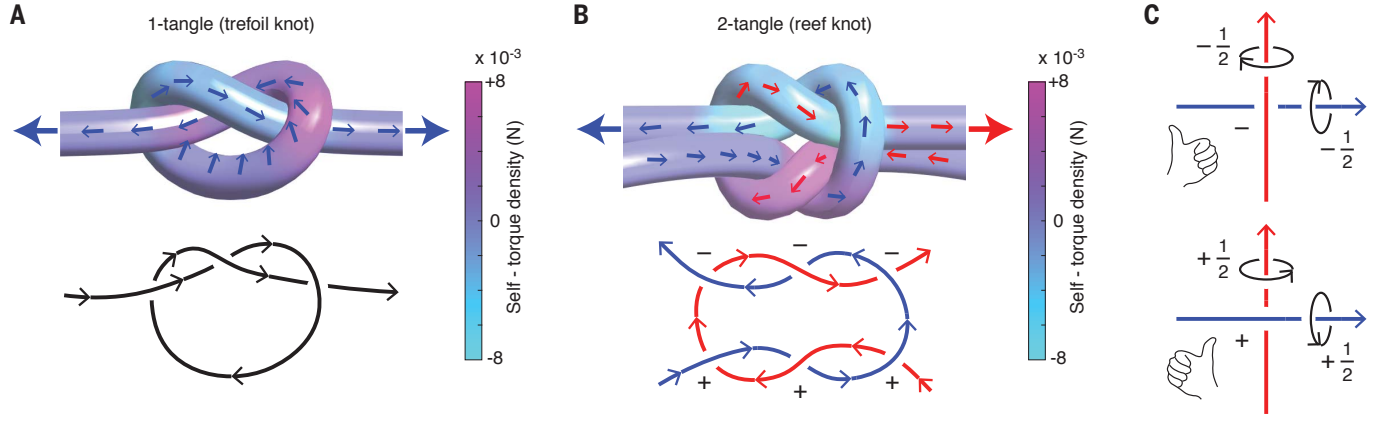


Fig. 2. Topology and self-twisting in 1-tangles and 2-tangles. (**A**) Top: A 1-tangle is tightened by pulling its two ends in opposite directions (large exterior arrows). The induced fiber velocity field (small interior arrows) in the center-of-mass frame reverses its orientation near the fiber midpoint. Bottom: As the velocity field is incompatible with any chosen global fiber orientation (black arrows), self-torque data cannot be consistently assigned to a topological 1-tangle diagram. (**B**) Top: Because of the presence of the two free ends, the pulling directions of a bend knot (large exterior arrows) define a canonical global orientation on each of the two fibers in the corresponding 2-tangle. Bottom: The alignment of local velocity directions and fiber orientation permits the discretization of self-torque data over crossings by assigning twist charges $q_i = \pm 1$ to each vertex i as described in the next panel. (**C**) Each individual fiber strand passing through vertex i induces a rotation in the other strand, thus contributing $\pm 1/2$ to the vertex twist charge $q_i = \pm 1$, with sign corresponding to rotation handedness. Blue–blue and red–red self-crossings found in more complex 2-tangles can be labeled accordingly. The sum of the q_i defines the total writhe Wr, providing a coarse-grained approximation of the total self-torque in 2-tangles; the reef knot has $Wr = 0$. Fiber diameter is 0.4 mm and pulling force is 15 N in (A) and (B).

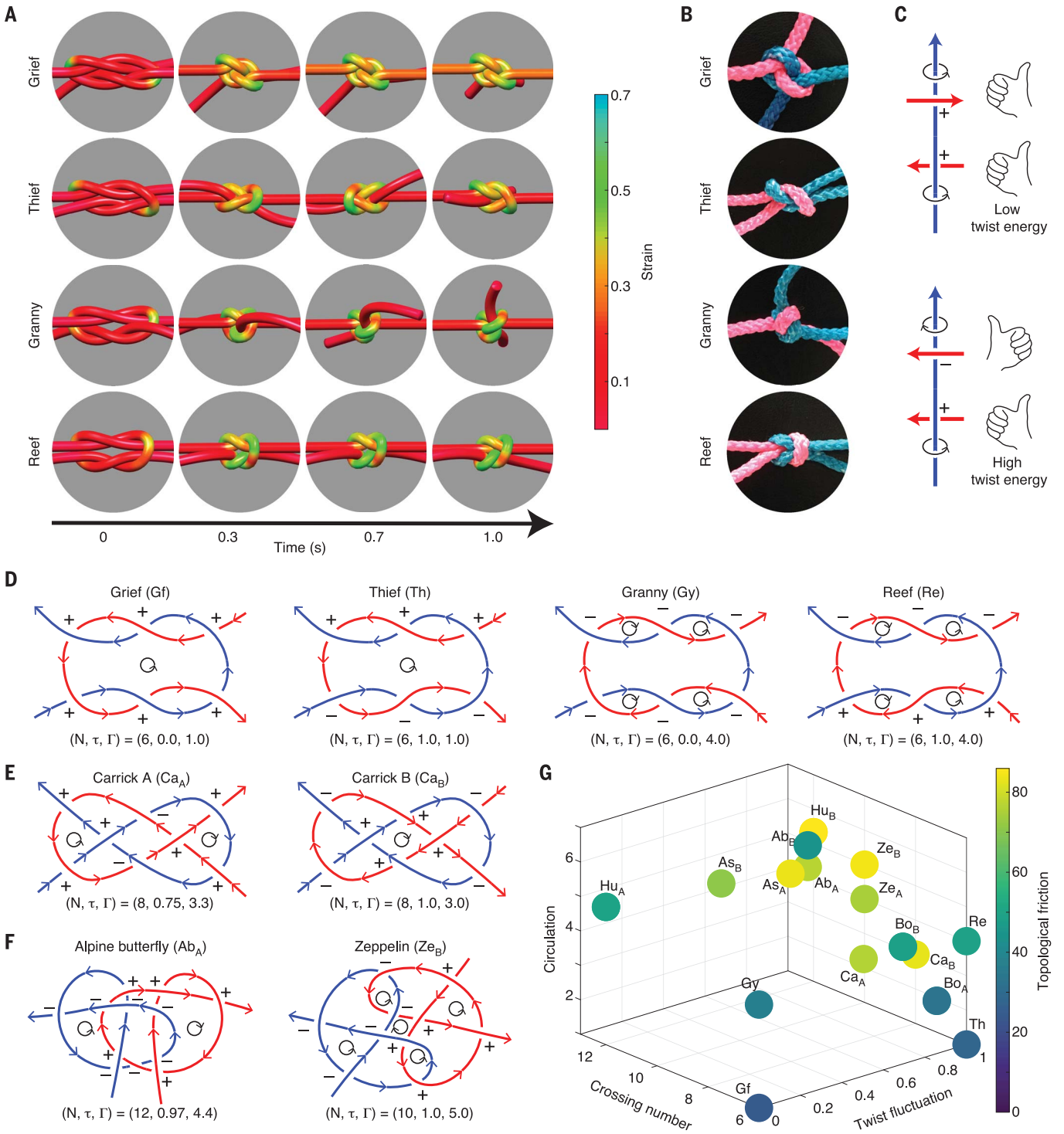


Fig. 3. Topology determines the mechanical stability of 2-tangles. (**A**) Small modifications in topology lead to substantial changes in the mechanical behavior of 2-tangles, exemplified by the presence or absence of global rotation of the knot body upon pulling (movie S2); fiber diameter is 0.4 mm and pulling force is 15 N. Knots are shown in order of least stable (grief knot) to most stable (reef knot). (**B**) Simulated tight configurations of knots are validated with real knots tied in nylon rope (diameter, 20 mm) with horizontal ends being pulled. (**C**) Tight knots act on themselves by right-handed (positive) and left-handed (negative) torques. Equally directed torques lead to rolling (top), whereas opposite torques promote locking (bottom) and thus stabilize a knot against untying. (**D** to **F**) Knot diagrams oriented by pulling direction correspond to a topological state defined as the triple of crossing number N, twist fluctuation τ, and circulation Γ. These parameters explain the relative stability of knots in the reef group (D) and the Carrick group (E). (F) The Zeppelin bend is more stable than the alpine butterfly bend, displaying both higher twist fluctuation and higher circulation. (**G**) Topological state reveals the underlying structure of bend knots and separates stable knots from unstable knots. The dimensionless topological friction, obtained from simulation, is determined by the velocity response when the knot is pulled with a given force and is a measure of the friction force caused by the knot (*28*). Labels in (G) correspond to those in (D), (E), and (F) and additional knots listed in fig. S3.

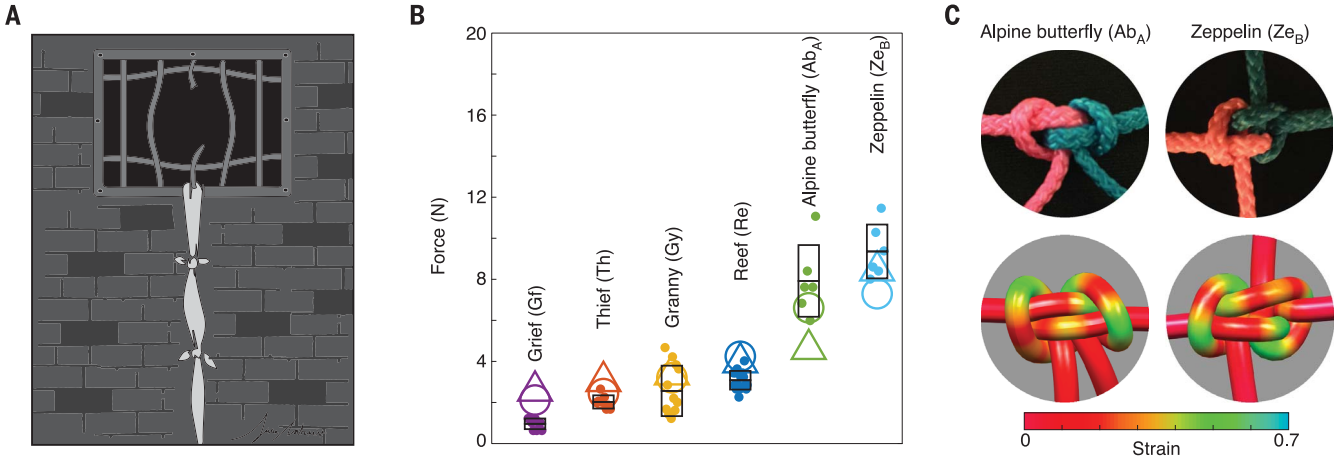


Fig. 4. Experiments for commonly used knots validate the theoretically predicted phase diagram. (**A**) Our experimental setup (*28*) mimics the classical prisoner's escape problem by determining the critical pulling force $F = mg$ at which two lines untie. (**B**) Experiments measuring the critical mass m at which two Dyneema fibers untie confirm the ranking of knot stability predicted by simulations. For simpler knots with crossing number ≤8, averages (horizontal lines) over individual experiments (small filled circles) agree quantitatively with the relative strength predicted from simulations measuring both the velocity-based friction (large empty circles) and the total compression force (large empty triangles) within the knot (*28*); black boxes indicate standard deviations of the individual experiments, with $N = 9$ grief (Gf), $N = 8$ thief (Th), $N = 12$ granny (Gy), $N = 16$ reef (Re), $N = 6$ alpine butterfly (Ab), and $N = 6$ Zeppelin (Ze) knots. For complex knots with high crossing number, such as the Zeppelin bend, more sophisticated models accounting for material-specific friction forces and three-dimensional contact details need to be developed in the future. Fiber diameter is 0.15 mm. (**C**) Nonetheless, simulations of complex bends with generic friction (*28*) show good shape agreement with tight configurations of bends in nylon ropes (diameter, 20 mm) and reveal the highly nonuniform strain distributions in such knots.

and C). In analogy to the coarse-graining procedure underlying Ising-type spin models, we can associate a unit twist charge $q_i = \pm 1$ with each vertex i in the planar 2-tangle diagram, where the sign of q_i reflects the combined handedness of the torques acting on the two intersecting strands (Fig. 2C). The sum $Wr = \sum_i q_i$, mathematically known as the writhe, represents the total self-torque of a 2-tangle, establishing a concrete link between topology and mechanics.

A key puzzle of physical knot theory (*9*), the empirically observed mechanical difference between the visually similar reef and granny knots (Fig. 3, A and B), may be understood as a consequence of this torque–writhe correspondence in 2-tangles. The underlying mechanism becomes evident by considering a pair of crossings as shown in Fig. 3C. Whereas equally directed torques lead to rolling, opposite torques promote locking and thus stabilize a knot against untying. The overall stability of 2-tangles therefore depends on the self-torque distribution along the fibers, as encoded by the vertex twist charges $q_i = \pm 1$ in an untightened knot diagram (Fig. 3, D to F). The above argument suggests the following topological twist fluctuation energy per site:

$$\tau = \frac{1}{N}\sum_i (q_i - \bar{q})^2 = \tau_0(N) - \frac{2}{N^2}\sum_{i<j} q_i q_j \qquad (1)$$

where N is the crossing number, $\bar{q} = (1/N)\sum_i q_i = Wr/N$ is the average writhe, and $\tau_0 = 1 - 1/N$ can be interpreted as a ground-state energy density (*28*). Equation 1 has the form of a ferromagnetic energy for an Ising-type spin model with long-range interactions, emphasizing the concept of knots as strongly coupled systems.

In addition to twist locking for large values of τ, knots can be stabilized when their internal structure forces fiber segments to slide tangentially against each other. For example, the reef knot and the thief knot both have $\tau = 1$, but because their pulled ends differ, friction makes the reef knot more stable (Fig. 3, D and G). At the coarse-grained level of planar knot diagrams, these friction effects correspond to edge-to-edge interactions dominated by pairs of edges sharing a face and pulled in opposite directions (Fig. 3, D to F). To formalize this notion, each edge around a face F is assigned a weight of +1 or −1 if it winds around F in the anticlockwise or clockwise direction, respectively. Each face then contributes a friction energy given by the net circulation of the edges around the face, C_F, normalized by the total number of edges e_F. This yields the total circulation energy:

$$\Gamma = \sum_F \frac{|C_F|}{e_F} \qquad (2)$$

where the sum is taken over all faces of the knot diagram. The normalization encodes the assumption that every face has the same perimeter in the tight limit, ensuring that each face contributes a maximum of +1 to Γ.

The topological parameters N, τ, and Γ allow us to rationalize the stability of a large class of popular knots used by sailors and climbers (Fig. 3G). These variables are easy to evaluate from knot diagrams (Fig. 3, D to F) and reflect topology-induced forces and torques throughout the knot. As such, the triplet (N, τ, Γ) captures both essential topological and mechanical structure hidden within knots. The (N, τ, Γ) phase diagram explains existing empirical knowledge for simple knots (*2*), as well as predictions of the Kirchhoff model about the relative strength and stability of more complex 2-tangles (Fig. 3G). We verified these predictions independently in experiments by mimicking the prisoner's escape problem (Fig. 4A) with two thin Dyneema fibers made from ultra–high molecular weight polyethylene tied together (*28*). Of the two pulled ends for each knot, one is fixed in the experimental apparatus and the other is perturbed while suspending incrementally higher masses until the knot pulls through. Although the Kirchhoff model cannot account for surface contact details (*30*), the experimental data for the critical loads agree quantitatively with the simulations for simple knots and, more importantly, confirm the predicted qualitative stability differences between various commonly used knots (Fig. 4B). Notably, both theory and experiments indicate that the Zeppelin knot is more secure against untying than the popular alpine butterfly knot (Fig. 4, B and C).

To conclude, the above analysis shows how basic topological counting rules can be used to estimate the relative stability of frequently encountered knots and tangles. From a theoretical perspective, the parallels with long-range coupled spin systems suggest that the statistical mechanics (*4*, *21*) of general knotted structures can be formulated within this framework. Tangled vortices (*12*, *17*) in complex fluids and defect loops in liquid crystals (*15*) may permit similar statistical descriptions through reduction to topological crossing diagrams. In elastic systems, joint experimental and theoretical progress is needed to untangle long-standing puzzles regarding the statistics of knots in DNA (*18*) and proteins (*20*, *21*), where thermal effects induce a finite persistence length, and other macroscopic structures (*8*, *24*). In sailing, climbing, and many other applications, nontopological material parameters and contact geometry (*30*) also play important mechanical roles and must be included in more refined continuum models to quantitatively describe practically relevant knotting phenomena. From a broader conceptual and practical perspective, the above topological mechanics framework seems well suited for designing and exploring new classes of knots with desired behaviors under applied load.

REFERENCES AND NOTES

1. J. C. Turner, P. van de Griend, Eds., *History and Science of Knots*, vol. 11 of *K & E Series on Knots and Everything* (World Scientific, 1996).
2. C. W. Ashley, *The Ashley Book of Knots* (Doubleday Books, 1944).
3. J. B. Trimbos, *Obstet. Gynecol.* **64**, 274–280 (1984).
4. C. C. Adams, *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots* (American Mathematical Society, 2004).
5. B. Audoly, N. Clauvelin, S. Neukirch, *Phys. Rev. Lett.* **99**, 164301 (2007).
6. M. K. Jawed, P. Dieleman, B. Audoly, P. M. Reis, *Phys. Rev. Lett.* **115**, 118302 (2015).
7. B. F. Bayman, *Am. J. Phys.* **45**, 185–190 (1977).
8. P. B. Warren, R. C. Ball, R. E. Goldstein, *Phys. Rev. Lett.* **120**, 158001 (2018).
9. J. A. Calvo, K. C. Millett, E. J. Rawdon, Eds., *Physical Knots: Knotting, Linking, and Folding Geometric Objects in R^3*, vol. 304 of *Contemporary Mathematics* (American Mathematical Society, 2001).
10. W. Thomson II, *Philos. Mag.* **34**, 15–24 (1867).
11. H. K. Moffatt, R. L. Ricca, "Helicity and the Călugăreanu invariant" in *Knots And Applications*, L. H. Kauffman, Ed. (World Scientific, 1995), pp. 251–269.
12. D. Kleckner, W. T. Irvine, *Nat. Phys.* **9**, 253–258 (2013).
13. M. W. Scheeler, W. M. van Rees, H. Kedia, D. Kleckner, W. T. M. Irvine, *Science* **357**, 487–491 (2017).
14. S. Kuei, A. M. Słowicka, M. L. Ekiel-Jeżewska, E. Wajnryb, H. A. Stone, *New J. Phys.* **17**, 053009 (2015).
15. U. Tkalec, M. Ravnik, S. Čopar, S. Žumer, I. Muševič, *Science* **333**, 62–65 (2011).
16. J. B. Taylor, *Phys. Rev. Lett.* **33**, 1139–1141 (1974).
17. D. Kleckner, L. H. Kauffman, W. T. Irvine, *Nat. Phys.* **12**, 650–655 (2016).
18. R. Stolz *et al.*, *Sci. Rep.* **7**, 12420 (2017).
19. A. R. Klotz, B. W. Soh, P. S. Doyle, *Phys. Rev. Lett.* **120**, 188003 (2018).
20. P. Virnau, L. A. Mirny, M. Kardar, *PLOS Comput. Biol.* **2**, e122 (2006).
21. R. C. Lua, A. Y. Grosberg, *PLOS Comput. Biol.* **2**, e45 (2006).
22. D. Goundaroulis *et al.*, *Polymers* **9**, 444 (2017).
23. R. C. Ball, M. Doi, S. F. Edwards, M. Warner, *Polymer* **22**, 1010–1018 (1981).
24. C. Baek, A. O. Sageman-Furnas, M. K. Jawed, P. M. Reis, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 75–80 (2018).
25. R. D. Kamien, *Eur. Phys. J. B* **1**, 1–4 (1998).
26. J. R. Goldman, L. H. Kauffman, *Adv. Appl. Math.* **18**, 300–332 (1997).
27. M. Kolle *et al.*, *Adv. Mater.* **25**, 2239–2245 (2013).
28. Materials and methods are available as supplementary materials.
29. M. Bergou, M. Wardetzky, S. Robinson, B. Audoly, E. Grinspun, *ACM Trans. Graph.* **27**, 63 (2008).
30. J. H. Maddocks, J. B. Keller, *SIAM J. Appl. Math.* **47**, 1185–1200 (1987).
31. V. P. Patil, Knot simulation code. Zenodo (2019); https://doi.org/10.5281/zenodo.3528928.

ACKNOWLEDGMENTS

J.D. thanks the Isaac Newton Institute for Mathematical Sciences for support and hospitality during the program "The Mathematical Design of New Materials" (supported by EPSRC grant EP/R014604/1) when work on this paper was undertaken. We thank J. Takagi for producing Fig. 4A. **Funding:** This work was supported by an Alfred P. Sloan Research Fellowship (J.D.), a Complex Systems Scholar Award from the James S. McDonnell Foundation (J.D.), the Brigham and Women's Hospital through a Stepping Strong Innovator Award (J.D.S. and M.K.), and the National Science Foundation through the "Designing Materials to Revolutionize and Engineer our Future" program (DMREF-1922321 to J.D.S. and M.K.). **Author contributions:** V.P.P. and J.D. developed theory. V.P.P. performed simulations, for which J.D.S. and M.K. provided data and code for converting strain into perceived knot color. J.D.S. and M.K. designed color-changing fibers and conceived optomechanical experiments. J.D.S. conducted optomechanical experiments and provided the description of the experiments in the supplementary materials. J.D.S. and V.P.P. performed in-lab prisoner's escape experiments. V.P.P. and J.D. wrote the first draft of the paper. All authors discussed and revised the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The code used for numerical simulations is available on Zenodo (*31*). All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6473/71/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S6
Tables S1 and S2
Movies S1 and S2
References (*32*–*33*)

5 August 2019; accepted 15 November 2019
10.1126/science.aaz0135

FLUORESCENT PROTEINS

Electrostatic control of photoisomerization pathways in proteins

Matthew G. Romei[1]*, Chi-Yun Lin[1]*, Irimpan I. Mathews[2], Steven G. Boxer[1]†

Rotation around a specific bond after photoexcitation is central to vision and emerging opportunities in optogenetics, super-resolution microscopy, and photoactive molecular devices. Competing roles for steric and electrostatic effects that govern bond-specific photoisomerization have been widely discussed, the latter originating from chromophore charge transfer upon excitation. We systematically altered the electrostatic properties of the green fluorescent protein chromophore in a photoswitchable variant, Dronpa2, using amber suppression to introduce electron-donating and electron-withdrawing groups to the phenolate ring. Through analysis of the absorption (color), fluorescence quantum yield, and energy barriers to ground- and excited-state isomerization, we evaluate the contributions of sterics and electrostatics quantitatively and demonstrate how electrostatic effects bias the pathway of chromophore photoisomerization, leading to a generalized framework to guide protein design.

Photoisomerizable chromophores, such as those in rhodopsins, phytochromes, photoactive yellow proteins, and fluorescent proteins (FPs), rotate around specific bonds after photoexcitation in the protein environment, which is essential to converting light energy into molecular motion (*1*). To investigate the role of the protein environment on tuning bound chromophore and/or ligand functionality, we chose to study FPs, a relatively simple model system consisting of an autocatalytically formed chromophore contained in a β barrel (*2*). The chromophore's local environment can markedly alter its photophysical properties, leading to a wide range of colors, fluorescence quantum yields (FQYs), and photoswitching characteristics (*3*). The chromophore's FQY increases by three orders of magnitude when contained in the protein scaffold compared with when it is free in solution (*4*). The dominant nonradiative decay process that lowers the chromophore's FQY is isomerization about either the phenolate (P) or imidazolinone (I) bonds, resulting in a P-ring flip or cis-trans isomerization, respectively (Fig. 1A) (*5*). This nonradiative decay process is enhanced in photoswitching FPs that are widely used for super-resolution microscopy (*3*). Modulating the probability between radiative and nonradiative decay, and for the latter, the propensity for P- or I-bond isomerization, epitomizes the essential features of protein control.

The most well-studied and intuitively appealing hypothesis for the chromophore's substantial increase in FQY in the protein suggests that steric confinement of the protein scaffold physically prevents the bond rotation required for nonradiative decay, as demonstrated by studies involving chemically locked or artificially confined chromophores (*6*). An alternative hypothesis identifies the role of electrostatics in modulating the FQY. After a perturbation to either the chromophore's electronic state (e.g., by photon absorption) or nuclear coordinates (e.g., by isomerization), a redistribution of the chromophore's electron density occurs, which is usually described as charge transfer between the rings. Consequently, the electric field exerted by the environment can either promote or hinder charge transfer and thus could control whether fluorescence or isomerization is more favorable after excitation (*7*).

In earlier work on split green fluorescent protein (GFP), we linked structure and function with energetics (*8*) and showed that the dominant energetic feature governing the competition between fluorescence and isomerization is the excited-state (ES) barrier for chromophore bond rotation (Fig. 1B, process 2). Here we present a systematic study investigating the contributions of sterics and electrostatics to energetic features of the chromophore's potential energy surface in both the ground state (GS) and ES. To experimentally probe these effects, we introduced a diverse range of substituents on the chromophore's P ring using amber suppression (*9*) with substituted tyrosine residues (*10*), taking advantage of the chromophore's autocatalytic maturation process (Fig. 2A). The electronic perturbation to the chromophore due to the substituent can be thought of as analogous to a perturbation of the protein environment around the chromophore that alters the chromophore's electronic properties, as suggested by past studies on polymethine dyes (*11*). As a model system, we chose the widely used photoswitchable FP Dronpa2 [the Met^{159}→Thr (M159T) mutant of Dronpa] because of the balance between its moderately high FQY and photoisomerization efficiency (*12*). We also include results from a nonphotoswitchable FP, a superfolder GFP construct, to generalize the scope of our conclusions.

We expressed wild-type and 10 Dronpa2 variants with chromophores containing electron-donating and electron-withdrawing substituents on the P ring (Fig. 2B). X-ray crystal structures confirm that the P-ring substituent(s) occupies a single orientation, except for the 3-F variant, which has two orientations (supplementary text S1 and fig. S2). Introduction of the substituent(s) causes little to no structural deviation compared with the wild type (supplementary text S2 and figs. S1 to S3). The absorption spectrum for each Dronpa2

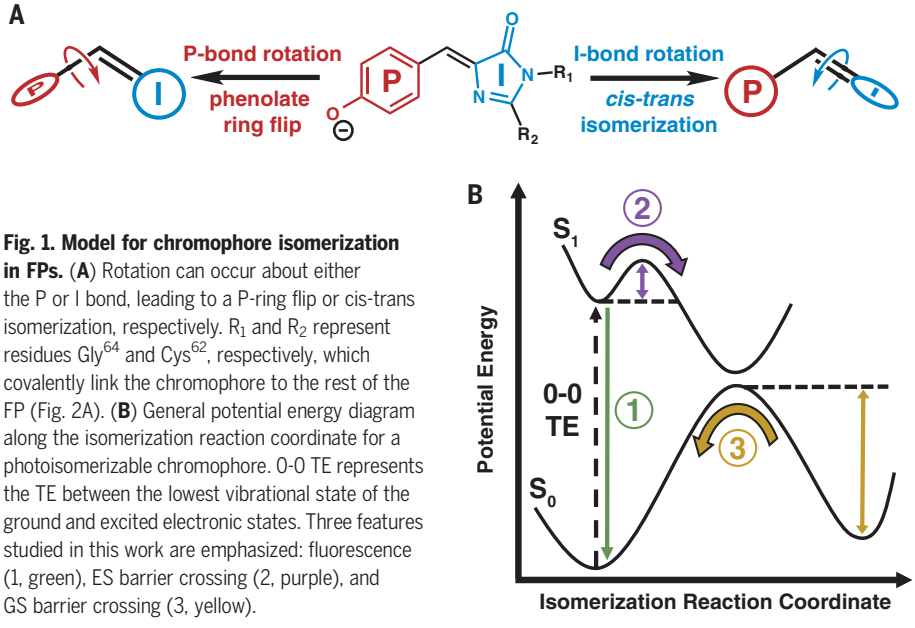


Fig. 1. Model for chromophore isomerization in FPs. (**A**) Rotation can occur about either the P or I bond, leading to a P-ring flip or cis-trans isomerization, respectively. R_1 and R_2 represent residues Gly^{64} and Cys^{62}, respectively, which covalently link the chromophore to the rest of the FP (Fig. 2A). (**B**) General potential energy diagram along the isomerization reaction coordinate for a photoisomerizable chromophore. 0-0 TE represents the TE between the lowest vibrational state of the ground and excited electronic states. Three features studied in this work are emphasized: fluorescence (1, green), ES barrier crossing (2, purple), and GS barrier crossing (3, yellow).

[1]Department of Chemistry, Stanford University, Stanford, CA 94305, USA. [2]Stanford Synchrotron Radiation Lightsource, Menlo Park, CA 94025, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: sboxer@stanford.edu

variant directly reflects the electronic contribution of the substituent: Electron-donating groups red-shift, whereas electron-withdrawing groups blue-shift the absorption maximum (Fig. 2C and fig. S4). Both the trend of electrostatic color tuning and the direction of charge transfer upon excitation can be understood through either a Hammett analysis (supplementary text S3) or Olsen's resonance color theory (*13*) (supplementary text S4 and fig. S8), agreeing with simulations that suggest negative charge flows from the P ring to the I ring upon photon absorption (*5*). We use the absorption peak maximum [an approximation of the 0-0 transition energy (TE)] (supplementary text S5) as a scale to reflect the substituents' electron-donating and electron-withdrawing capabilities (supplementary text S4) because the initial Franck-Condon excitation is purely an electronic process.

We then sought to examine the influence of each chromophore's steric and electronic properties on ES processes. First, we measured each variant's FQY (Fig. 1B, process 1) and plotted the values against the corresponding TE (Fig. 3A). The trend is nonmonotonic with a peaked shape; variants with red- and blue-shifted TEs show positive and negative correlations with FQY, respectively. A variant of superfolder GFP with the same series of substituted chromophores exhibits the same trend (fig. S7), suggesting that electrostatic sensitivity is an intrinsic chromophore property.

To elucidate the underlying energetics, we estimated the ES energy barrier (Fig. 1B, process 2) for each Dronpa2 variant using temperature-dependent fluorescence lifetime measurements (supplementary texts S6 and S7 and figs. S10 to S15), which capture the combined decay rate of all relaxation processes from the S_1 minimum. As with FQY, the ES barrier heights show a peaked trend when plotted against TE (Fig. 3B). Linear fits to the electron-donating and electron-withdrawing variants' data exhibit slopes with similar magnitude but opposite sign (Fig. 3D), which describe the extent and direction of charge transfer during the ES barrier crossing. A change of 1 kcal/mol in TE in either direction corresponds to a change of ~1.5 kcal/mol in ES barrier height, implying that ES barrier crossing, a nonradiative process, is more sensitive to electronic effects than Franck-Condon excitation.

To investigate the role of steric and electronic effects on GS isomerization barrier height (Fig. 1B, process 3), we determined the isomerization rate constant through pH-dependent thermal relaxation kinetics measurements after photoexcitation to a cis-trans photostationary state, assuming the validity of transition state theory (supplementary text S8 and figs. S17 to S18). A plot of the GS barrier height versus TE (fig. S19) appears to show a lack of correlation

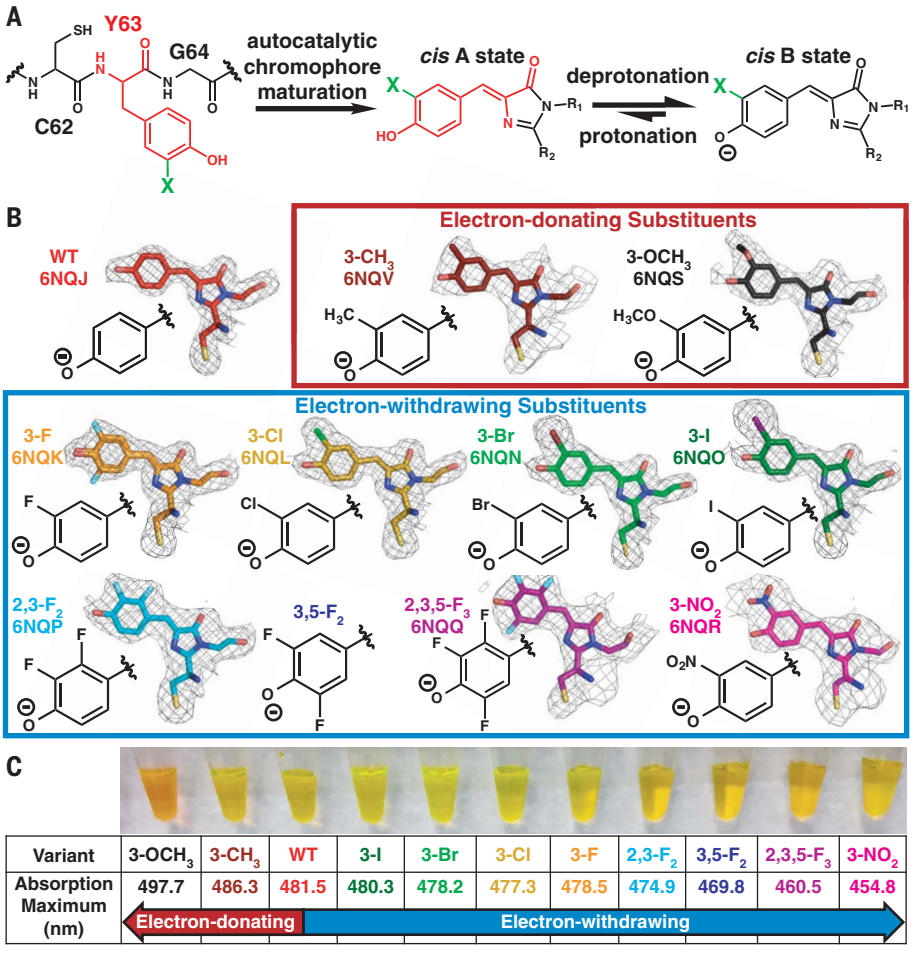
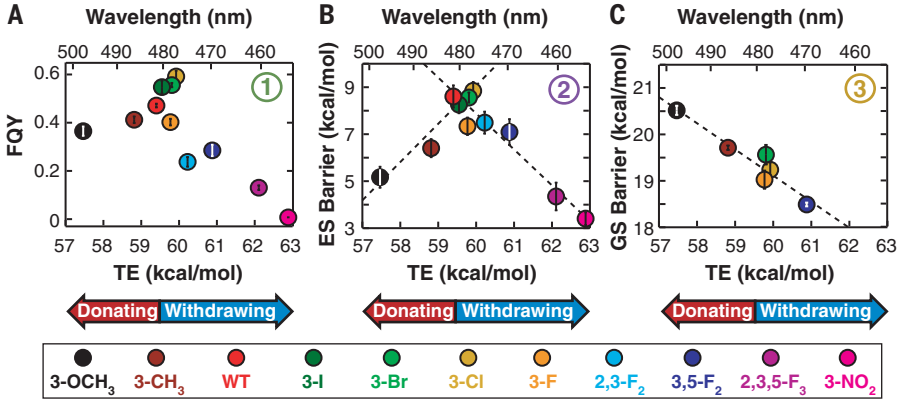


Fig. 2. Incorporation of electron-donating and electron-withdrawing substituents into the Dronpa2 chromophore. (**A**) Scheme depicting incorporation of substituents (represented by a green "X") through amber suppression of Y63 and chromophore maturation in Dronpa2 variants (C, Cys; Y, Tyr; G, Gly). (**B**) Dronpa2 amber suppression variants grouped by electron-donating and electron-withdrawing properties. The electron density maps ($2mF_o - DF_c$, 1σ) from solved x-ray structures (except 3,5-F_2, which could not be crystallized; see supplementary text S2) show substituent orientation(s) (see fig. S2 for omit maps). Two conformations were necessary for modeling the chromophore of the 3-F variant (fig. S2). The legend of fig. S2 includes the identity of the monomer displayed for each variant. WT, wild type. (**C**) Image of purified proteins and their corresponding 77 K absorption peak maxima.

between the GS barrier and the substituent's electronic effects. Close examination of fig. S19 reveals that the substituent's steric properties may also contribute to the observed trends. For example, among the data points for the 3-F, 3-Cl, 3-Br, and 3-I substituents, as highlighted by the gray box in fig. S19, the barrier height increases as a function of halogen size despite similar TEs, indicating that substituent size influences GS barrier height. To isolate the electrostatic contribution to GS barrier height, we created an isosteric substituent series (defined in supplementary text S9) and plotted the corresponding data for this subgroup, which monotonically decrease as a function of TE (Fig. 3C). The extent of charge transfer during GS barrier crossing reflected by the slope in Fig. 3C is approximately one-third of that in the ES (Fig. 3B and charge-transfer extent in Fig. 3D), suggesting that changes in the electronic properties of the chromophore have a smaller, but still evident, impact on thermal relaxation. In contrast to the GS barrier, the influence of sterics on the ES barrier is minimal (Fig. 3D and fig. S20). If sterics were the dominant factor, large substituents would be expected to increase the barrier to chromophore twisting in the ES and, consequently, FQY. However, electrostatics is clearly the dominant factor for ES isomerization in a constant protein environment (Fig. 3, A and B, and fig. S20).

The observation of two approximately equal but opposite slopes between ES barrier height and TE (Fig. 3B) suggests a mechanism change for barrier crossing that depends on the electronic properties of the chromophore (supplementary text S3). Charge transfer is coupled



D **Summary of ground and excited state properties of Dronpa2 amber suppression variants**

Photophysical Process		Charge Transfer Extent [a]	Electron Flow	Sensitivity to Sterics
Excitation		+1 (by definition)	P ⟶ I	no
ES Barrier Crossing	Withdrawing	-1.5	I ⟶ P	minimal
	Donating	+1.6	P ⟶ I	minimal
GS Barrier Crossing		-0.6	I ⟶ P	yes

[a] Refers to the degree of charge transfer of a given process, determined by the slopes in B and C, compared to that of the excitation process

Fig. 3. GS and ES properties of Dronpa2 amber suppression variants. Circled numbers refer to processes depicted in Fig. 1B. (**A**) FQY versus TE and (**B**) ES energy barrier height versus TE both exhibit a peaked shape (see fig. S7 for comparable FQY results from GFP variants). Linear fits to the electron-donating and electron-withdrawing substituent data in (B) are shown as dashed lines with positive and negative slope, respectively. (**C**) Isosteric series of the GS energy barrier height plotted against TE, with a dashed line representing a linear fit. (**D**) Summary of GS and ES properties of Dronpa2 variants.

with bond twisting in the ES of monomethine dyes (*5*, *14*). In the GFP chromophore, this twisting can proceed about either the P or I bond (Fig. 1A). Both P and I twisting in the isolated anionic chromophore have comparable barrier heights but opposite charge-transfer directions, suggesting that small perturbations to the chromophore or its environment could influence which twisting pathway is more energetically favorable (*14*). Our data reveal a decrease in FQY and ES barrier height compared with those from wild-type Dronpa2 regardless of the substituent's electronic effect (Fig. 3, A and B), indicating that the electronic properties of the chromophore can tune the relative ES barrier heights for P and I twisting, thus biasing toward the pathway with the lower barrier. By contrast, the GS chromophore follows the same isomerization mechanism regardless of the substituent's electronic effect (Fig. 3C).

The observed trends in energetics can be linked to structural intuition through a valence-bond model depicting the deprotonated chromophore as an allylic anion (Fig. 4, A and B) (*5*, *14*, *15*). In Fig. 4C, we present an energetic model for chromophore isomerization guided by the adiabatic states in Fig. 4B. The electronic properties of the chromophore, governed in this work by the substituent on the P ring, determine the ES barrier height difference between P and I twisting. For GS

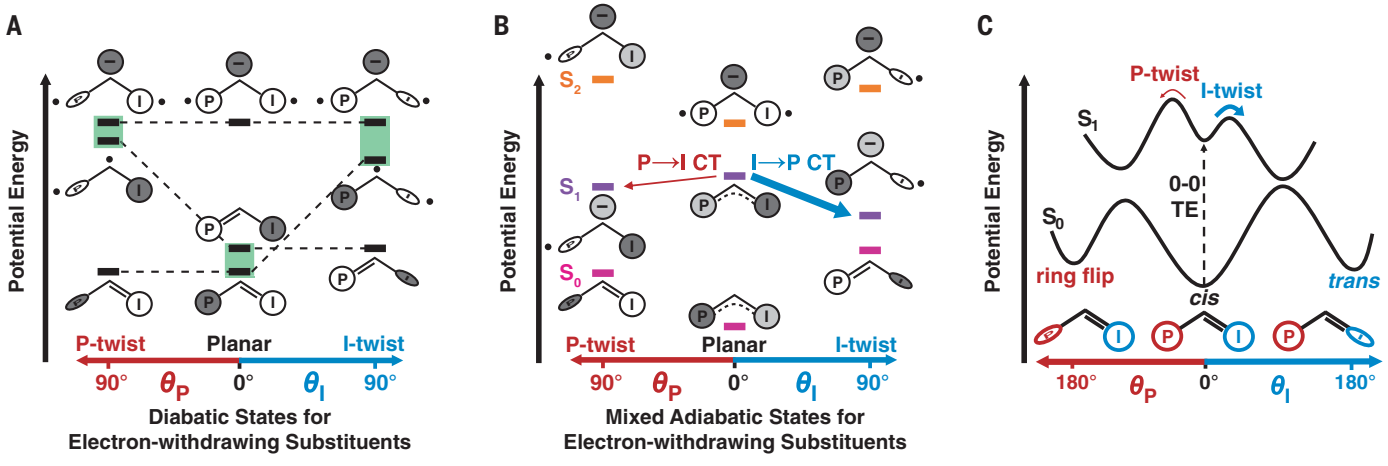


Fig. 4. Allylic anion model of isomerization for a chromophore containing an electron-withdrawing substituent. Shades of gray represent relative magnitude of negative charge localized to the methine bridge, P ring, or I ring, and dots represent unpaired electrons. The color scheme is equivalent to that in Fig. 1A. (**A**) Three diabatic states of the chromophore in planar, I-twisted, and P-twisted geometries, with energetic penalties required for breaking double bonds for rotation. Mixing of the coupled states (highlighted in green) leads to the adiabatic states shown in (**B**). For variants with electron-withdrawing substituents, the I-twist pathway is more energetically downhill, and thus preferred, compared with the P-twist pathway. Electron-donating substituents would have the opposite energetic effect and favor the P-twist pathway (not shown for clarity). Although the relative energy levels of this allylic anion model are qualitative, they are consistent with high-level calculations on the free chromophore at different bond rotation geometries (*15*). Negative charge transfer (CT) occurs from I to P for the I-twist pathway and from P to I for the P-twist pathway, which agrees with a Hammett analysis (supplementary text S3) and simulations of the free chromophore (*5*, *14*). (**C**) Potential energy diagram for FP chromophore isomerization with two competing bond rotation pathways inspired by the mixed adiabatic states in (B). The GS cis chromophore is excited from S_0 to S_1 and relaxes to an S_1 local minimum (relaxation coordinate not shown) (*13*). From the S_1 minimum, the chromophore rotates about either the P or the I bond, depending on the relative ES barrier heights of the competing processes. The diagram represents Dronpa2 variants with electron-withdrawing substituents; variants with electron-donating substituents would have an inverted barrier height ratio between the two competing twisting pathways.

relaxation from trans to cis, we are limited by what we can measure spectroscopically, namely the photochromic cis-trans isomerization. We can therefore observe only one GS twisting mechanism (I twist), which explains the monotonic trend in GS barrier height as a function of TE (Fig. 3C). In the potential energy diagram (Fig. 4C), the transition state for either the P- or I-twist pathway in S_1 (ES) lies closer to planarity than the corresponding transition state in S_0 (GS). As a result, reaching the transition state on the GS surface requires greater bond rotation, explaining the enhanced steric sensitivity observed for GS isomerization.

By engineering chromophore variants using amber suppression, we have systematically elucidated the role of electrostatics on chromophore color and isomerization in an FP environment. The electrostatic sensitivities of the chromophore stem from the intrinsic direction of charge transfer during electronic transitions and photoisomerizable bond rotations, which is ubiquitous in other photoisomerizable systems (*8*, *16*–*24*). By tuning the environment of the chromophore in these protein systems, with an emphasis on the often-overlooked electrostatic component, it may be possible to finely control properties of interest, such as regioselective isomerization, because of distinctive charge redistributions as different bonds are rotated. On the basis of our results in FPs, introducing hydrogen-bond–donating residues around the P ring of the chromophore would bias toward the I-twist photoisomerization pathway (*13*). In the photoisomerizable retinal chromophore in rhodopsins, theoretical studies have suggested that different bond-specific photoisomerization intermediates have different electronic distributions (*21*), allowing for similar targeted environmental modifications to bias bond rotation pathways. As such, our conclusions provide an initial, generalizable framework to incorporate electrostatic and steric effects into the design of other photoisomerizable systems to help develop improved variants and new functionalities in optogenetics and imaging (*1*).

REFERENCES AND NOTES

1. C. P. O'Banion, D. S. Lawrence, *ChemBioChem* **19**, 1201–1216 (2018).
2. R. Y. Tsien, *Annu. Rev. Biochem.* **67**, 509–544 (1998).
3. D. Bourgeois, V. Adam, *IUBMB Life* **64**, 482–491 (2012).
4. A. Acharya *et al.*, *Chem. Rev.* **117**, 758–795 (2017).
5. M. E. Martin, F. Negri, M. Olivucci, *J. Am. Chem. Soc.* **126**, 5452–5464 (2004).
6. C. L. Walker *et al.*, *Curr. Opin. Chem. Biol.* **27**, 64–74 (2015).
7. J. W. Park, Y. M. Rhee, *J. Am. Chem. Soc.* **138**, 13619–13629 (2016).
8. C.-Y. Lin, J. Both, K. Do, S. G. Boxer, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E2146–E2155 (2017).
9. Amber suppression allows for the site-specific incorporation of noncanonical amino acids into recombinant proteins. The amino acid site of interest on the relevant gene is mutated to the amber stop codon, TAG. Non-native tRNA containing an anticodon recognizing the amber stop codon is encoded on a separate plasmid. A non-native aminoacyl tRNA synthetase is also introduced into the cell that is capable of charging the non-native tRNA with the noncanonical amino acid. The tRNA then recognizes the amber stop codon on the mRNA within the translation machinery of the cell, allowing for the incorporation of the noncanonical amino acid into the growing polypeptide chain.
10. A. Dumas, L. Lercher, C. D. Spicer, B. G. Davis, *Chem. Sci.* **6**, 50–69 (2015).
11. S. R. Marder *et al.*, *Science* **265**, 632–635 (1994).
12. A. C. Stiel *et al.*, *Biochem. J.* **402**, 35–42 (2007).
13. C.-Y. Lin, M. G. Romei, L. M. Oltrogge, I. I. Mathews, S. G. Boxer, *J. Am. Chem. Soc.* **141**, 15250–15265 (2019).
14. S. Olsen, K. Lamothe, T. J. Martínez, *J. Am. Chem. Soc.* **132**, 1192–1193 (2010).
15. S. Olsen, R. H. McKenzie, *J. Chem. Phys.* **130**, 184302 (2009).
16. S. Schenkl, F. van Mourik, G. van der Zwan, S. Haacke, M. Chergui, *Science* **309**, 917–920 (2005).
17. C. Ko, A. M. Virshup, T. J. Martínez, *Chem. Phys. Lett.* **460**, 272–277 (2008).
18. L. S. Wolfe *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 16863–16868 (2010).
19. S. Gozem, I. Schapiro, N. Ferré, M. Olivucci, *Science* **337**, 1225–1228 (2012).
20. G. Bassolino *et al.*, *J. Am. Chem. Soc.* **136**, 2650–2658 (2014).
21. S. Gozem, H. L. Luk, I. Schapiro, M. Olivucci, *Chem. Rev.* **117**, 13502–13565 (2017).
22. D. Smyrnova, M. d. C. Marín, M. Olivucci, A. Ceulemans, *J. Chem. Theory Comput.* **14**, 3163–3172 (2018).
23. M. d. C. Marín *et al.*, *J. Am. Chem. Soc.* **141**, 262–271 (2019).
24. C. Punwong, S. Hannongbua, T. J. Martínez, *J. Phys. Chem. B* **123**, 4850–4857 (2019).

ACKNOWLEDGMENTS

We dedicate this manuscript to the memory of Seth Olsen, whose theoretical studies of the GFP chromophore motivated much of the analysis of this work. We thank S. Lynch of the Stanford NMR Facility for assistance with NMR data collection and interpretation. R. A. Mehl of the Unnatural Protein Facility was instrumental in providing an aminoacyl-tRNA synthetase for 3-methyltyrosine incorporation. S. H. Schneider helped develop the MATLAB code for statistical analysis of the fluorescence lifetime data. We thank J. I. Brauman, T. J. Martínez, N. H. List, S. D. Fried, and L. M. Oltrogge for discussions regarding this work. **Funding:** M.G.R. was supported by a Center for Molecular Analysis and Design graduate fellowship. C.-Y.L. was supported by a Kenneth and Nina Tai Stanford Graduate Fellowship and the Taiwanese Ministry of Education. This work was supported, in part, by the NIH (grant GM118044 to S.G.B.). Part of this work was performed at the Stanford Nano Shared Facilities (SNSF) and supported by the National Science Foundation (award ECCS-1542152). Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences (contract no. DE-AC02-76SF00515). The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by the National Institutes of Health (NIH), National Institute of General Medical Sciences (NIGMS) (including P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the NIGMS or NIH. **Author contributions:** M.G.R. and C.-Y.L. designed the experiments, expressed the proteins, performed the experiments, interpreted the results, and wrote the manuscript. I.I.M. assisted with protein crystallization, collected x-ray diffraction data, and assisted with data refinement. S.G.B. discussed results and wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All x-ray density maps and atomic models for Dronpa2 variants have been deposited in the Protein Data Bank (wild type: 6NQJ; 3-F: 6NQK; 3-Cl: 6NQL; 3-Br: 6NQN; 3-I: 6NQO; 2,3-F_2: 6NQP; 2,3,5-F_3: 6NQQ; 3-NO_2: 6NQR; 3-CH_3: 6NQV; 3-OCH_3: 6NQS). All other data are presented in the main text or supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6473/76/suppl/DC1
Materials and Methods
Supplementary Texts S1 to S9
Figs. S1 to S27
Tables S1 to S11
References (*25*–*86*)

View/request a protocol for this paper from *Bio-protocol*.

1 March 2019; resubmitted 4 June 2019
Accepted 31 October 2019
10.1126/science.aax1898

OPTICS

On-chip integrated laser-driven particle accelerator

Neil V. Sapra[1]*, Ki Youl Yang[1], Dries Vercruysse[1], Kenneth J. Leedle[1], Dylan S. Black[1], R. Joel England[2], Logan Su[1], Rahul Trivedi[1], Yu Miao[1], Olav Solgaard[1], Robert L. Byer[1], Jelena Vučković[1]

Particle accelerators represent an indispensable tool in science and industry. However, the size and cost of conventional radio-frequency accelerators limit the utility and reach of this technology. Dielectric laser accelerators (DLAs) provide a compact and cost-effective solution to this problem by driving accelerator nanostructures with visible or near-infrared pulsed lasers, resulting in a 10^4 reduction of scale. Current implementations of DLAs rely on free-space lasers directly incident on the accelerating structures, limiting the scalability and integrability of this technology. We present an experimental demonstration of a waveguide-integrated DLA that was designed using a photonic inverse-design approach. By comparing the measured electron energy spectra with particle-tracking simulations, we infer a maximum energy gain of 0.915 kilo–electron volts over 30 micrometers, corresponding to an acceleration gradient of 30.5 mega–electron volts per meter. On-chip acceleration provides the possibility for a completely integrated mega–electron volt-scale DLA.

Dielectric laser accelerators (DLAs) have emerged as a promising alternative to conventional radio-frequency accelerators because of the large damage threshold of dielectric materials (*1*, *2*); the commercial availability of powerful, near-infrared femtosecond pulsed lasers; and the low-cost, high-yield nanofabrication processes that produce them. Together, these advantages allow DLAs to make an impact in the development of applications requiring mega–electron volt energy beams of nanoampere currents, such as tabletop free-electron lasers, targeted cancer therapies, and compact imaging sources (*3*–*7*).

DLAs are designed by choosing an appropriate pitch and depth of a periodic structure such that the near fields are phase matched to electrons of a specific velocity (*8*, *9*). These structures, together with focusing elements, integrated electron sources, and microbunching structures, form the building blocks to achieve mega–electron volt-scale energy gain through cascaded stages of acceleration (*10*–*13*). Previous demonstrations of DLAs have relied on free-space lasers directly incident on the accelerating structure, often pillars or gratings made of fused silica or silicon (*14*–*20*). However, free-space excitation requires bulky optics; therefore, integration with photonic circuits would enable increased scalability, robustness, and impact of this technology.

Integration with photonic waveguides represents a design challenge because of difficulties in accounting for scattering and reflections of the waveguide mode from subwavelength features. Although tuning the geometric parameters and location of a few etched holes in the waveguide is possible (*21*), this requires brute-force optimization of only a small subset of the design space. Instead, we used an inverse-design approach to develop a waveguide-integrated DLA on a 500-nm device layer silicon-on-insulator (SOI) platform, which allows for expansion of the design space (*22*). This on-chip accelerator is demonstrated by coupling light from a pulsed laser through a broadband grating coupler and exciting a waveguide mode that acts as the source for the accelerator (Fig. 1A).

To meet the phase-matching condition, the periodicity of the accelerating structure, Λ, is set by $\Lambda = \beta\lambda$, where $\beta = v/c$ is the ratio of the velocities of the incident electrons to the speed of light and λ is the center wavelength of the pump laser (*23*). To match experimental parameters, we designed for a center pump wavelength of 2 μm and an input electron velocity of $v = 0.5c$, resulting in an accelerator period of $\Lambda = 1$ μm. Fig. 1B captures the geometry of the optimization problem. Using

[1]E. L. Ginzton Laboratory, Stanford University, Stanford, CA, USA. [2]SLAC National Accelerator Laboratory, Menlo Park, CA, USA.
*Corresponding author. Email: nvsapra@stanford.edu

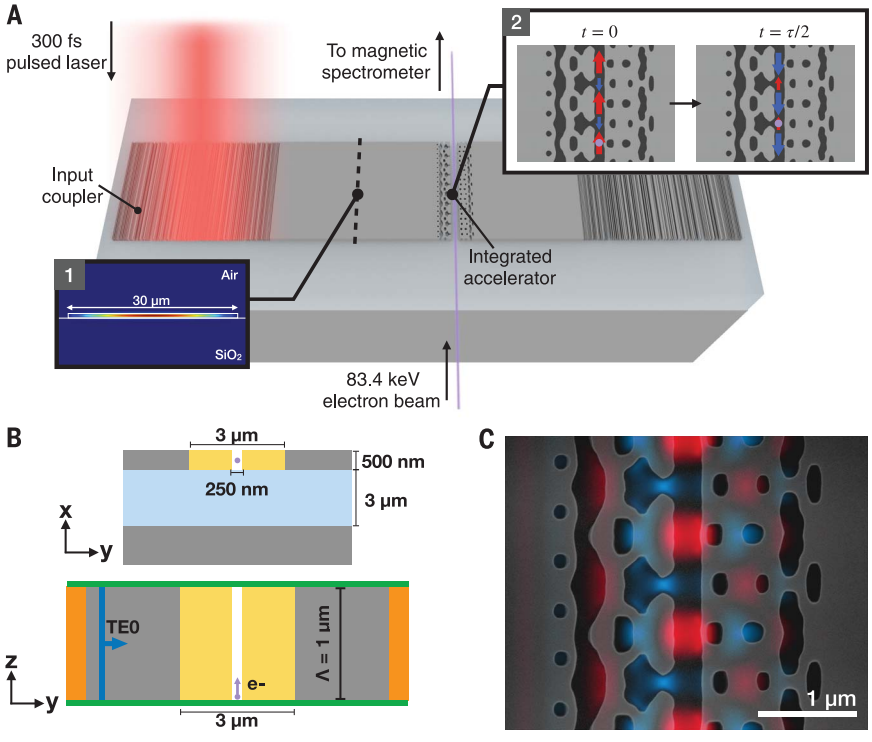


Fig. 1. Inverse design of on-chip particle accelerator. (**A**) Schematic (not to scale) depicting components of the on-chip accelerator. An inverse-designed grating couples light from a normally incident free-space beam into the fundamental mode of a slab waveguide (inset 1). The excited waveguide mode then acts as the excitation source for the accelerating structure. The accelerator structure, also created through inverse design, produces near fields that are phase matched to an input electron beam with initial energy of 83.4 keV. Inset 2 depicts the phase-matched fields and electron at half an optical cycle ($\tau/2$) apart. (**B**) Geometry of the optimization problem. We designed on a 500-nm silicon (gray), 3-μm buried oxide layer (light-blue) SOI material stack. Periodic boundary conditions (green) are applied in the *z*-direction, with a period of $\Lambda = 1$ μm, and perfectly matched layers were used in the remaining directions (orange). We optimized the device over a 3-μm design region (yellow) with an input source of the fundamental TE0 mode. During the optimization, a 250-nm channel for the electron beam to travel in is maintained. (**C**) SEM image of the final accelerator design obtained from the inverse-design method. A frame from a time-domain simulation of the accelerating fields, E_z, is overlaid.

an in-house inverse-design software suite (*24*–*27*), the design of the accelerator was optimized over a 3-μm region, ensuring the preservation of a 250-nm center channel for electron propagation. The accelerator was simulated with a fully-3D finite-difference frequency-domain (FDFD) solver meshed with a uniform grid spacing of 30 nm. Periodic boundary conditions were applied in the direction of electron propagation (z-axis) to enforce the accelerator period, and perfectly matched layers were used in the remaining axes (*28*). The structure was excited with the fundamental slab waveguide mode and the following 3D optimization problem was solved:

$$\underset{p,E^1,E^2,\ldots,E^m}{\text{maximize}} \sum_{i=1}^{m} |G_z(E^i)| - |G_y(E^i)|$$
$$\text{subject to } \nabla \times \frac{1}{\mu_0} \nabla \times E^i - \omega_i^2 \epsilon(p) E^i = -i\omega_i J^i, \quad i = 1, 2, \ldots, m \tag{1}$$

We expressed the acceleration gradient, $G_z(E^i)$, the integrated field that the electron experiences as it travels through one period of the accelerator, in the frequency domain (*29*). The second term, $G_y(E^i)$, corresponds to the deflecting transverse gradients, which we penalized. The fields were subjected to Maxwell's equations and the permittivity of the device, $\epsilon(p)$, was parameterized by a vector of design variables, p (*30*). This vector describes the permittivity of the device during the first, continuous stage of optimization and a level-set function that defines the boundaries of the device in the second, discrete stage of optimization. To have good spectral overlap with the broadband input pulsed laser spectrum, each objective function evaluation is the sum of $m = 3$ simulations, each with a different input source frequency, ω_i. The three simulations uniformly sample a 30-nm total bandwidth around 2 μm. During the final optimization stage, an additional constraint was introduced to enforce a minimum fabricable feature size of 80 nm. Further details regarding the design of the accelerator and a time-lapse movie of the optimization can be found in the supplementary materials (*31*). A scanning electron microscope (SEM) image of a fabricated optimized accelerator is shown with a frame from simulated time-domain fields overlaid (Fig. 1C).

As the optimization was performed with periodic boundary conditions, the performance of a finite-length 30-period accelerator structure was verified in a 3D finite-difference time-domain (FDTD) simulation (*32*). The frequency response of the grating coupler and accelerator were computed and cascaded to determine the frequency-domain acceleration gradient (*31*). This complete acceleration gradient spectrum is shown in Fig. 2. The spectrum peaks at $\lambda = 1.964$ μm (Fig. 2A), indicating a shift from the design parameters caused by the finite length of the structure and numerical dispersion (*33*). With knowledge of the peak operating wavelength, the time-domain characteristics were modeled by propagating a 300-fs unchirped Gaussian pulse, centered at 1.964 μm, through the grating coupler, waveguide, and accelerator (*31*). The time-domain acceleration gradient (G_z) and deflecting gradients (G_y, G_x) are given by:

$$G_k(t_0) = \frac{1}{L}\int_0^L E_k(z, t_0 + z/\beta c_0)dz \tag{2}$$

where t_0 is the delay between the time of source injection and the electron entering the accelerator channel and $L = 30$ μm is the length of the accelerator. The accelerating and deflecting gradients down the center of the channel are evaluated at t_0, which maximizes the acceleration gradient, $G_z(t_0)$ (Fig. 2B). As the time-domain gradients are normalized to the peak incident pulse amplitude, Fig. 2B also provides the simulated structure factors, the ratio of acceleration gradient to incident field. Although we obtained good suppression of the deflecting gradients, one can also operate at another time delay, t_0', such that the deflecting gradients, $G_y(t_0')$ and $G_x(t_0')$, are further minimized in the center of the channel.

A 30-period accelerator, waveguides, and grating couplers were fabricated on a 500-nm-thick SOI wafer using electron beam lithography and reactive ion etching. The input grating coupler was separated by 50 μm of waveguide from the accelerator structure, and the output coupler was separated by 30 μm of waveguide from the accelerator. The entire structure had a width of 30 μm. To provide clearance for the electron beam, the area surrounding the accelerator was etched with an additional photolithography step to form a "mesa" (Fig. 3). Complete fabrication details can be found in the materials and methods section (*31*).

The experimental setup was adapted from previous direct-incidence pillar experiments to support normal incidence on a grating coupler (*13*, *18*). Light, polarized in the direction of electron propagation (z), generated from a 300-fs FWHM pulse-length, 100-kHz repetition rate optical parametric amplifier was focused to a 40-μm, $1/e^2$-diameter spot. The beam is normally incident on the input grating coupler to excite the fundamental TE0 waveguide mode of the slab waveguide [for grating coupler design, see the materials and methods (*31*)]. A custom-built scanning transmission electron microscope was used as the source for the electron beam, which travels through the channel in the accelerator structure with an initial energy of 83.4 keV ($v = 0.51c$). Electrons that passed through the accelerator were separated by energy in a magnetic spectrometer before terminating at a microchannel plate detector to image the energy distribution [see the materials and methods section for additional details (*31*)].

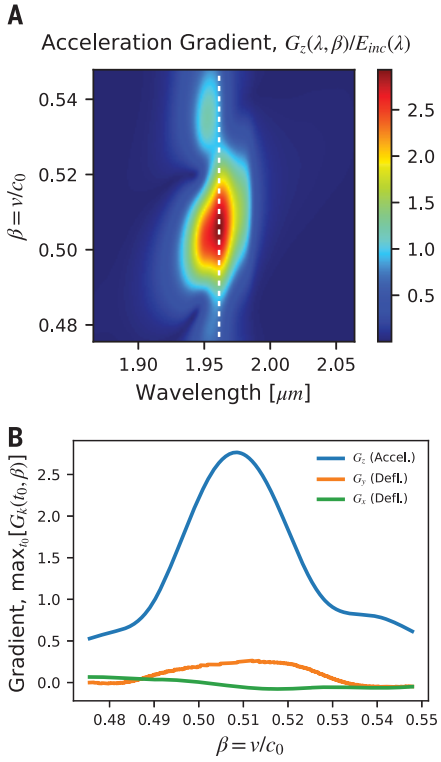


Fig. 2. Simulated performance of optimized accelerators. (**A**) Acceleration gradient spectrum for a finite-length accelerator composed of 30 periods, including frequency response of grating coupler. The gradient is normalized to the maximum frequency-domain amplitude of the incident Gaussian beam. Dashed line indicates optimal operation wavelength of the simulated structure, $\lambda = 1.964$ μm. (**B**) Time-domain accelerating gradients and transverse deflecting gradients as a function of input electron energy from simulated fields. Gradients were evaluated at time-delay, t_0, which maximizes the acceleration gradient, G_z. Fields normalized to the peak electric field of the pulse incident on the grating coupler.

The electron energy spectra (Fig. 4A) showed that electrons had been successfully accelerated by our structure. The blue curve depicts the energy spectrum of the electrons passing through the accelerator structure with the laser off, and the red curve shows the energy spectrum when the laser (3 mW average power, 335 MV/m peak field, at 1.94 μm) was incident on the grating coupler. Because the bunch length was larger than the optical cycle, we observed symmetric broadening of the energy spectrum, resulting in electrons being accelerated and decelerated. To characterize the broadening of the laser-on spectra, we introduced an energy spectrum width metric, ξ, which we define as the first trailing energy

at which the difference between the laser-on spectra $F(\varepsilon)$ and the laser-off spectra $f(\varepsilon)$ was <0.01: $F(\xi) - f(\xi) \leq 0.01$. For the spectra shown in Fig. 4A, centered at ε_0 = 83.4 keV, this corresponds to a value of ξ = 84.31 keV. The dotted red curve depicts simulated performance of the accelerator based on a commercial particle-tracking code to propagate a distribution of particles consistent with experimental parameters through the 3D electromagnetic field map of Fig. 1C, providing agreement with the experimental spectrum (*31*). Because of the spread in energy and phase of the input electron spectrum, the maximal energy gain is a quantity not directly measurable from the laser-on spectrum. Instead, we could obtain this value from the particle-tracking simulations (*31*). From these simulations, we inferred a maximal energy gain of 0.915 keV over 30 μm, providing a gradient of 30.5 MeV/m and a structure factor (ratio of acceleration gradient to incident field) of 0.09.

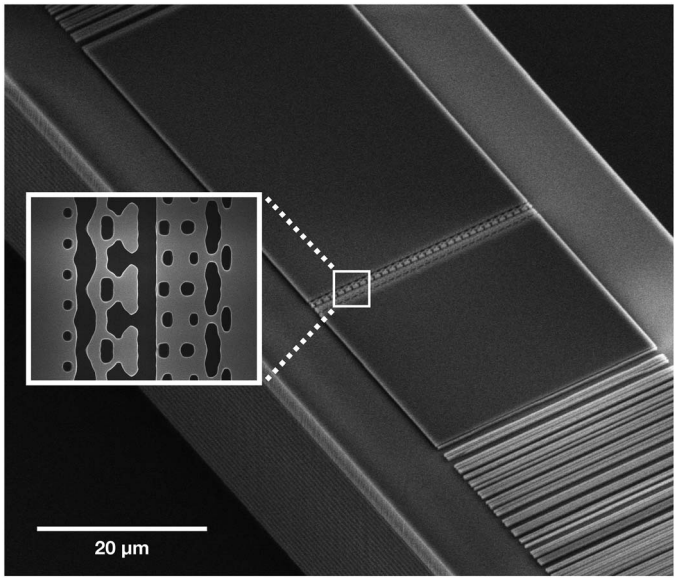


Fig. 3. Fabricated single-stage accelerator. SEM image of a single-stage accelerator of 30 periods fabricated on a 500-nm SOI stack. The accelerator sits on a 25-μm-tall mesa structure to provide clearance for the input electron beam.

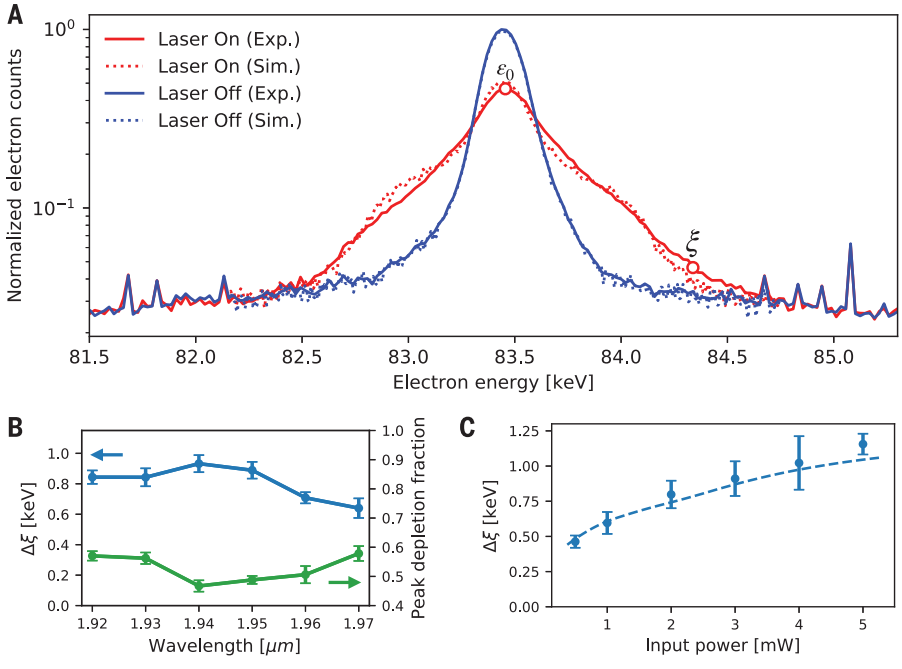


Fig. 4. Experimental verification of accelerator. (**A**) Electron energy spectrum (log-scale) without laser incident (blue curve) and with laser incident (3.0 mW, 335 MV/m peak field, at λ = 1.94 μm; red curve) on the grating coupler. Simulated spectrum is based on particle-tracking simulations shown in the dotted red curve. On the spectra, ε_0 denotes the center energy of the distribution and ξ provides an energy spectral width metric that marks the energy at which the difference between the laser-on and laser-off spectra is below 0.01. (**B**) Energy spectral width broadening, $\Delta\xi = \xi - \varepsilon_0$, (blue, left axis), and peak depletion (green, right axis) for a fixed power at 2.75 mW, 321 MV/m peak field, as a function of varying the wavelength of the pump laser. (**C**) Measured energy spectral width, $\Delta\xi$, at a fixed wavelength of 1.94 μm as a function of input power, with simulation from the tracking code superimposed as a dashed curve.

To determine the operating wavelength of our accelerator, the average power of the incident laser pulses on the grating coupler were fixed to be 2.75 mW (321 MV/m peak field) and the wavelength was swept (Fig. 4B). A peak was observed in the energy spectrum width, $\Delta\xi = \xi - \varepsilon_0$, at 1.94 μm. Moreover, the ratio of laser-on to laser-off counts at the center energy, referred to as "peak depletion," was optimal at 1.94 μm, consistent with an increase in the number of modulated electrons at this wavelength. The greatest broadening of the energy spectra and dip in peak depletion suggested an operating wavelength of 1.94 μm. This wavelength was blue-shifted from the simulated operating value of 1.964 μm. Because of the cavity-like nature of this accelerator, we attribute this spectral shift and flattening of the gradient spectrum to fabrication imperfections. Additionally, as a consequence of the blue shift, the $\Lambda = \beta\lambda$ design condition was no longer satisfied exactly and so some dephasing was to be expected, contributing to the diminished structure factor. Fixing the wavelength to 1.94 μm (Fig. 4C), we conducted a sweep of the input power from 0.5 to 5 mW (137 to 433 MV/m peak fields). As we increased the power, the measured values of the spectral width, $\Delta\xi$, compared favorably with those obtained from the simulated particle-tracking spectra indicated by the dashed curve in Fig. 4C (*31*).

Although nonlinear dephasing has been observed in other DLA experiments (*34*, *35*), the short waveguide distances (50 μm) in this experiment were much smaller than the hundreds of micrometers of propagation distance required to introduce nonlinear dephasing (*36*). Additionally, coupling into higher-order modes of the slab waveguide, specifically the TE2 waveguide mode, can result in dephasing. However, with 95% of the total power in the TE0 mode and only 3.8% of power expected to couple into TE2, this negative contribution would be minimal (*31*). Although not catastrophic to operation of the accelerator, postexperiment SEM imaging revealed laser-induced damage at the input grating coupler. Additional characterization identified this damage to occur after 3 to 4 mW of input power. System-level analysis of an SOI integrated accelerator such as the one presented here predicts acceleration gradients of 45.3 MeV/m (*36*). This suggests that work

toward higher-efficiency grating couplers with greater resiliency to high-field hotspots provides an achievable path to this estimated value.

The fabricated devices accelerate electrons of an initial energy of 83.4 keV by an inferred maximum energy gain of 0.915 keV over 30 μm, demonstrating acceleration gradients of 30.5 MeV/m. In this integrated form, these devices, alongside focusing and bunching elements, can be cascaded to reach mega–electron volt-scale energies capitalizing on the inherent scalability of photonic circuits.

REFERENCES AND NOTES

1. B. C. Stuart, M. D. Feit, A. M. Rubenchik, B. W. Shore, M. D. Perry, *Phys. Rev. Lett.* **74**, 2248–2251 (1995).
2. A.-C. Tien, S. Backus, H. Kapteyn, M. Murnane, G. Mourou, *Phys. Rev. Lett.* **82**, 3883–3886 (1999).
3. P. G. O'Shea, H. P. Freund, *Science* **292**, 1853–1858 (2001).
4. R. W. Hamm, M. E. Hamm, *Industrial Accelerators and Their Applications* (World Scientific, 2012).
5. S. Hanna, *RF Linear Accelerators for Medical and Industrial Applications* (Artech House, 2012).
6. K. Wootton, J. McNeur, K. Leedle, *Reviews of Accelerator Science and Technology: Volume 9: Technology and Applications of Advanced Accelerator Concepts* (World Scientific, 2016), pp. 105–126.
7. T. Plettner, R. Byer, *Phys. Rev. Spec. Top. Accel. Beams* **11**, 030704 (2008).
8. R. J. England *et al.*, *Rev. Mod. Phys.* **86**, 1337–1389 (2014).
9. T. Plettner, P. Lu, R. Byer, *Phys. Rev. Spec. Top. Accel. Beams* **9**, 111301 (2006).
10. J. McNeur *et al.*, *Optica* **5**, 687 (2018).
11. U. Niedermayer, T. Egenolf, O. Boine-Frankenheim, P. Hommelhoff, *Phys. Rev. Lett.* **121**, 214801 (2018).
12. B. Naranjo, A. Valloni, S. Putterman, J. B. Rosenzweig, *Phys. Rev. Lett.* **109**, 164803 (2012).
13. D. S. Black *et al.*, *Phys. Rev. Lett.* **122**, 104801 (2019).
14. E. A. Peralta *et al.*, *Nature* **503**, 91–94 (2013).
15. J. Breuer, P. Hommelhoff, *Phys. Rev. Lett.* **111**, 134803 (2013).
16. K. J. Leedle *et al.*, *Opt. Lett.* **40**, 4344–4347 (2015).
17. K. P. Wootton *et al.*, *Opt. Lett.* **41**, 2696–2699 (2016).
18. K. J. Leedle *et al.*, *Opt. Lett.* **43**, 2181–2184 (2018).
19. D. Cesar *et al.*, *Opt. Express* **26**, 29216–29224 (2018).
20. Z. Chen, K. Koyama, M. Uesaka, M. Yoshida, R. Zhang, *Appl. Phys. Lett.* **112**, 034102 (2018).
21. C.-M. Chang, O. Solgaard, *Appl. Phys. Lett.* **104**, 184102 (2014).
22. S. Molesky *et al.*, *Nat. Photonics* **12**, 659–670 (2018).
23. J. Breuer, R. Graf, A. Apolonski, P. Hommelhoff, *Phys. Rev. Spec. Top. Accel. Beams* **17**, 021301 (2014).
24. J. Vuckovic *et al.*, *SPINS-B: Inverse Design Software for Nanophotonic Structures, Stanford OTL Reference S18-012* (2018); https://github.com/stanfordnqp/spins-b.
25. A. Y. Piggott *et al.*, *Nat. Photonics* **9**, 374–377 (2015).
26. A. Y. Piggott, J. Petykiewicz, L. Su, J. Vučković, *Sci. Rep.* **7**, 1786 (2017).
27. L. Su, A. Y. Piggott, N. V. Sapra, J. Petykiewicz, J. Vuckovic, *ACS Photonics* **5**, 301–305 (2017).
28. W. Shin, S. Fan, *J. Comput. Phys.* **231**, 3406–3431 (2012).
29. T. Hughes, G. Veronis, K. P. Wootton, R. Joel England, S. Fan, *Opt. Express* **25**, 15414–15427 (2017).
30. D. Vercruysse, N. V. Sapra, L. Su, R. Trivedi, J. Vučković, *Sci. Rep.* **9**, 8999 (2019).
31. See the supplementary materials for additional information.
32. FDTD: 3D Electromagnetic Simulator (Lumerical Inc., 2019); https://www.lumerical.com/products/.
33. A. Taflove, S. C. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method* (Artech House, 2005).
34. D. Cesar *et al.*, *Commun. Phys.* **1**, 46 (2018).
35. S. Tan *et al.*, *Opt. Lett.* **44**, 335–338 (2019).
36. T. W. Hughes *et al.*, *Phys. Rev. Appl.* **9**, 054017 (2018).

ACKNOWLEDGMENTS

Funding: We thank all the members of the Accelerator on a Chip International Program for discussion and collaboration. This work was supported by the Gordon and Betty Moore Foundation (grant no. GBMF4744) and the U.S. Department of Energy, Office of Science (grant nos. DE-AC02-76SF00515 and DE-SC0009914). K.Y.Y. acknowledges funding from a Nano- and Quantum Science and Engineering Postdoctoral Fellowship. D.V. acknowledges funding from FWO and the European Union Horizon 2020 Research and Innovation Program (under Marie Sklodowska-Curie grant no. 665501). R.T. acknowledges a Kailath Graduate Fellowship. Part of this work was performed at the Stanford Nano Shared Facilities (SNSF)/Stanford Nanofabrication Facility (SNF), which is supported by the National Science Foundation under award no. ECCS-1542152.
Author contributions: N.V.S. performed and led the design, simulation, and fabrication of the accelerator. K.Y.Y. and Y.M. assisted with fabrication. D.V. assisted with design. K.J.L. and D.S.B. conducted the electron acceleration experiment. R.J.E. performed the particle-tracking simulations. L.S. provided the grating coupler design code. R.T. assisted in simulation analysis. J.V., R.L.B., and O.S. organized the collaboration and supervised the experiments. All authors participated in the discussion and interpretation of the results. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6473/79/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S4
References (*37*–*39*)
Movies S1 and S2

5 July 2019; accepted 26 November 2019
10.1126/science.aay5734

NEUROSCIENCE

Dendritic action potentials and computation in human layer 2/3 cortical neurons

Albert Gidon[1], Timothy Adam Zolnik[1], Pawel Fidzinski[2,3], Felix Bolduan[4], Athanasia Papoutsi[5], Panayiota Poirazi[5], Martin Holtkamp[2], Imre Vida[3,4], Matthew Evan Larkum[1,3]*

The active electrical properties of dendrites shape neuronal input and output and are fundamental to brain function. However, our knowledge of active dendrites has been almost entirely acquired from studies of rodents. In this work, we investigated the dendrites of layer 2 and 3 (L2/3) pyramidal neurons of the human cerebral cortex ex vivo. In these neurons, we discovered a class of calcium-mediated dendritic action potentials (dCaAPs) whose waveform and effects on neuronal output have not been previously described. In contrast to typical all-or-none action potentials, dCaAPs were graded; their amplitudes were maximal for threshold-level stimuli but dampened for stronger stimuli. These dCaAPs enabled the dendrites of individual human neocortical pyramidal neurons to classify linearly nonseparable inputs—a computation conventionally thought to require multilayered networks.

The expansion of the human brain during evolution led to an extraordinarily thick cortex (~3 mm), which is disproportionately thickened in layers 2 and 3 (L2/3) (*1*). Consequently, human cortical neurons of L2/3 constitute large and elaborate dendritic trees (*2*, *3*), decorated by numerous synaptic inputs (*1*). The active electrical properties of these dendrites largely determine the repertoire of transformations of the synaptic inputs to axonal action potentials (APs) at the output. Thus, they constitute a key element of the neuron's computational power.

We used dual somatodendritic patch clamp recordings and two-photon imaging to directly investigate the active properties of L2/3 dendrites in acute slices from surgically resected brain tissue of the human neocortex from epilepsy and tumor patients. Subthreshold (steady-state) potentials attenuated from the dendrite to the soma with a length constant (λ_{steady}) of 195 μm (fig. S1; n = 23 cells). In the opposite direction, the back-propagating action potentials (bAPs) attenuated from the soma to the dendrite with a λ_{bAP} of 290 μm (Fig. 1, A to C; n = 31 cells). Both λ_{bAP} and λ_{steady} were shorter than the length of the apical dendrite (the somata of these cells were located ~850 μm below the pia mater, on average, and the apical dendrite extended up to layer 1), which implies that strong attenuation governs the electrical activity to and from most synapses located on the apical dendrite.

We filled cells with the calcium indicator Oregon-green BAPTA-1 (100 μM) and measured the change in fluorescence ($\Delta F/F$) under a two-photon microscope while triggering APs at the soma. Trains of somatic APs (10 APs) at 50 Hz failed to cause Ca^{2+} influx at distal apical dendrites (fig. S2). AP trains with a higher frequency (10 APs at 200 Hz) did invade most of the apical dendrite, similarly to what has been shown previously in rodent L2/3 pyramidal neurons (*4*). However, these high-frequency signals were substantially attenuated at distal tuft dendrites (fig. S2). Furthermore, Ca^{2+} influx in spines was similar to that in the nearby dendritic branches, regardless of the somatic AP frequency (fig. S2D).

We next examined whether human L2/3 dendrites have intrinsic mechanisms to compensate for the large dendritic attenuation. We injected a current step into the dendrite (I_{dend}) and recorded the membrane potentials at both the dendrite and at the soma. At the soma and at the proximal dendritic sites (170 μm from the soma, on average), a suprathreshold current readily evoked somatic APs, which back-propagated into the dendrite (Fig. 1, A and B, and fig. S5H). However, when the dendritic electrode was positioned more distally, suprathreshold stimuli often evoked trains of repetitive APs that were initiated exclusively in the dendrite (Fig. 1D; for transient stimulus, see fig. S10). These results imply that L2/3 dendrites in human cortical pyramidal neurons are distinctly more excitable than the homolog dendrites in rodents, where similar steady currents evoke, at most, only a single dendritic AP at the beginning of the voltage response (*5*). In contrast to L2/3 pyramidal neurons, layer 5 pyramidal neurons of the human neocortex were recently reported to have reduced dendritic excitability compared with their homolog neurons in rodents (*6*).

High-frequency dendritic APs (>200 Hz) that were uncoupled from somatic firing have been observed in rodent dendrites in vivo (*7*, *8*). The authors of these studies have attributed these spikes to dendritic voltage-gated Na^+ channels and/or *N*-methyl-D-aspartate (NMDA) receptors. The dendritic APs in human L2/3 neurons were not blocked by the sodium channel blocker tetrodotoxin (1 μM; n = 4 cells; fig. S3), but they were abolished by the Ca^{2+} channel blocker Cd^{2+} (200 μM; n = 5 cells; fig. S3). The dendritic Ca^{2+} APs that we observed in human L2/3 neurons have not been described in the cortical neurons of other mammalian species. Dendritic APs that are mediated (or are assumed to be mediated) by sodium currents in rodents' neurons have been variously named dendritic spikes (*9*), prepotentials (*10*), Na-dSpikes (*11*), and dendritic action potentials (DAPs) (*8*). To distinguish the dendritic APs that we found in the human dendrites from those described previously, we refer to them as dendritic Ca^{2+} APs (dCaAPs).

dCaAPs were present not only in neurons from the temporal lobe of epilepsy patients but also in neurons from other neocortical areas of tumor patients (n = 4 cells from 3 patients; fig. S4). This suggests that dCaAPs are neither regionally confined nor related to pathology.

The waveform of dCaAPs was stereotypical and easily distinguished from that of bAPs. dCaAPs were typically wider than bAPs (with widths of 4.4 ± 1.4 ms, ranging between 2.6 and 8.0 ms; n = 32 cells), they were slow rising, and they did not have a kink at onset (*7*) (Fig. 1D). The majority of the cells (27 of 39) showed a train of (two or more) dCaAPs with a mean firing rate of 4.6 ± 1.7 Hz (dCaAPs per second). In the remaining 12 dendrites, a single dCaAP was triggered immediately after the beginning of the stimulus. Unlike the bAP (Fig. 1C), the amplitude of the dCaAPs (Fig. 1E) and their upstroke (fig. S5) were not dependent on the distance from the soma (average dCaAP amplitude 43.8 ± 13.8 mV, ranging between 13.0 and 67.0 mV; n = 32 cells, measured at threshold). This is consistent with both variability of the dCaAP initiation site and variability of dCaAP properties (for further details, see figs. S5 and S11). We never detected high-amplitude, long-duration, Ca^{2+} mediated plateau potentials, which are common in the apical dendrites of L5 neurons in rodents.

The impact of dCaAPs on the soma was variable. In some of the cells (17 of 37), the dCaAPs were coupled with somatic APs (coupled dCaAPs; e.g., Fig. 1F). Unlike forward-propagating dendritic APs in other pyramidal neurons (*12*–*14*), coupled dCaAPs triggered somatic APs immediately and/or with a delay ranging between 21.6 and 116.9 ms (53.8 ± 26.8 ms, on average, in 11 out of 17 coupled cells; Fig. 1, F and G, and fig. S6). Coupled dCaAPs that triggered somatic APs with a delay were classified as complex.

[1]Institute for Biology, Humboldt-Universität zu Berlin, Berlin, Germany. [2]Epilepsy-Center Berlin-Brandenburg, Department of Neurology, Charité - Universitätsmedizin Berlin, Berlin, Germany. [3]NeuroCure Cluster, Charité - Universitätsmedizin Berlin, Berlin, Germany. [4]Institute of Integrative Neuroanatomy, Charité-Universitätsmedizin Berlin, Berlin, Germany. [5]Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology - Hellas (IMBB-FORTH), Crete, Greece.
*Corresponding author. Email: matthew.larkum@hu-berlin.de

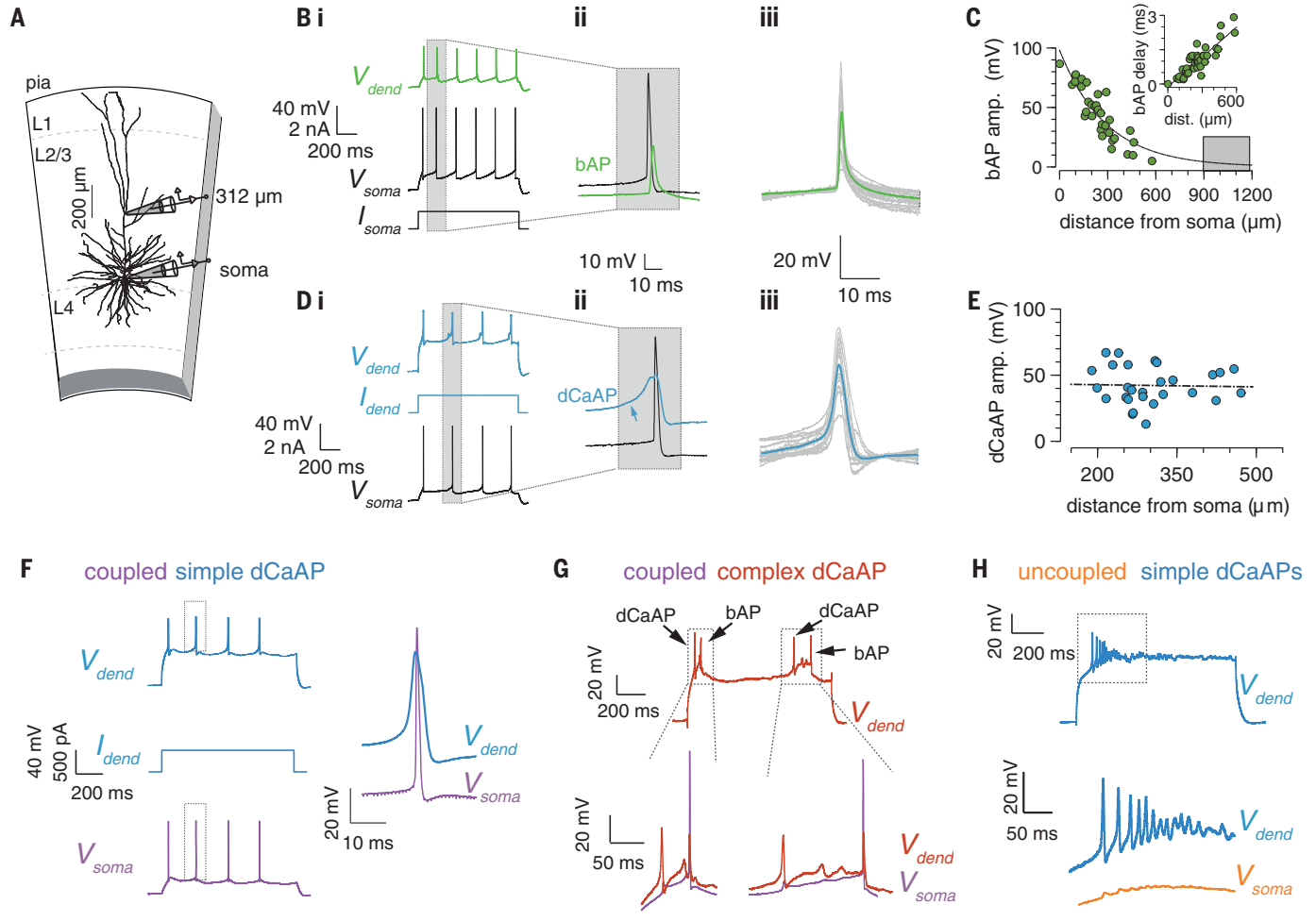


Fig. 1. bAPs and dCaAPs in human dendrites of L2/3 neurons. (**A**) Experimental setting: L2/3 neuron at a depth of 1130 µm below the pial surface, with a somatic electrode and a dendritic electrode placed at 312 µm from the soma. (**B**) (i) Recordings from the cell in (A). bAPs in green (V_{dend}) and corresponding somatic APs in black (V_{soma}) triggered by somatic current injection (I_{soma}) are shown. (ii) Somatic AP preceded the bAP [magnified from the frame in (i)]. (iii) bAPs in 16 dendrites (gray) and their average (green) aligned to their peak. (**C**) bAP amplitude (green dots) and exponential fit (length constant λ_{bAP} = 290 µm; n = 31 cells; dashed line) against distance from the soma. Gray area indicates the putative tuft region in layer 1 for the longest dendrite. (Inset) Delay of the bAP peak against distance from the soma with linear fit (r^2 = 0.78, where r^2 is the coefficient of determination). amp., amplitude; dist., distance. (**D**) (i) dCaAPs (V_{dend}; blue) triggered by a square current injected at the dendrite (I_{dend}) and the resulting somatic AP (black) from the cell in (A). (ii) Somatic AP (in black) and a dCaAP (in blue) magnified from (i). The slow rising dCaAP (blue arrow) precedes the somatic AP. (iii) Initial dCaAP in each recording at threshold in the same 16 dendrites (gray) in (iii) of (B) and their average trace (blue) aligned to their peak. (**E**) dCaAP amplitude is independent of the distance from the soma (n = 28 cells). Linear fit is shown with the dashed line (r^2 = 0.0009). (**F**) (Left) Coupled and simple dCaAPs (blue trace) and somatic APs (purple trace) triggered by I_{dend}. (Right) Magnified dCaAP (in blue) and a somatic AP (in purple) framed in the traces on the left are shown. (**G**) (Top) Two coupled and complex dCaAPs (in red) triggered delayed somatic APs [in purple, magnified at (bottom)]. (**H**) Burst of simple and uncoupled dCaAPs in blue (top) with somatic APs (bottom).

Without exception, each coupled dCaAP triggered a single somatic AP, which implies that, unlike calcium APs in the dendrites of other neurons (*15*, *16*), dCaAPs did not induce bursts of somatic APs. In the other 20 cells, dCaAPs were uncoupled. They were confined to the apical dendrite, unable to evoke somatic APs. Typically, uncoupled dCaAPs were observed in more distal dendritic recording sites (335 ± 113 µm from the soma) than the coupled dCaAPs that triggered somatic APs (265 ± 71 µm from the soma), but the distance difference was not statistically significant (Wilcoxon rank sum test, P = 0.077). Additionally, one coupled and three uncoupled cells fired bursts of three or more dCaAPs at the beginning of the stimulus (28 to 73 Hz). In fig. S7, we summarize the classification of the dCaAPs on the basis of their ability to trigger APs at the soma (i.e., coupled versus uncoupled) and their complexity (i.e., complex and/or simple). Most of the dendrites with complex dCaAPs also triggered simple dCaAPs, suggesting that their behavior might be activity- or input-dependent and/or modulated by other factors (*17*).

dCaAPs affected the input–output transformation of the cells. Typically, somatic AP firing increases with the input current intensity injected to the soma. In contrast, in 4 cells (out of 12 cells that had repetitive and coupled dCaAPs) our recordings revealed an inverse behavior where increasing the intensity of dendritic (rather than somatic) current injection resulted in decreased somatic firing. For example, in Fig. 2, A and B, the dendritic electrode evoked one or two somatic APs with current near threshold but failed to evoke APs for higher current intensity. In contrast, at the

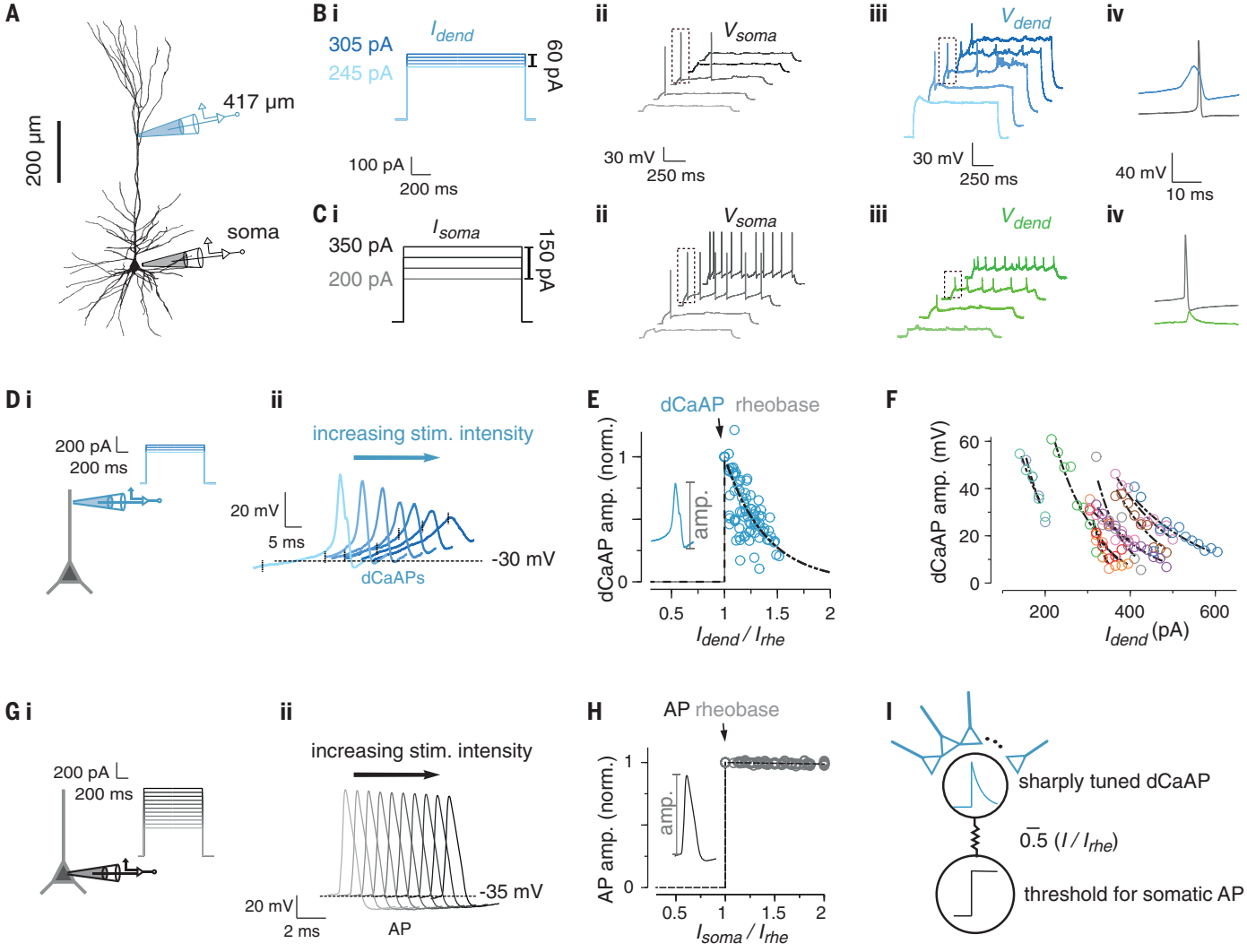


Fig. 2. dCaAPs are sharply tuned to the stimulus intensity. (**A**) L2/3 pyramidal neuron with soma 886 µm below the pia. The somatic and dendritic electrodes are shown in black and blue, respectively. Recordings from this cell are shown in (B) and (C). (**B**) Dendritic current (I_{dend}) injected 417 µm from the soma (i) and corresponding somatic (ii) and dendritic traces (iii). (ii) I_{dend} of 260 and 275 pA, but neither smaller nor larger current, resulted in somatic APs. (iii) dCaAP amplitudes were maximal for I_{dend} of 260 and 275 pA, whereas I_{dend} > 275 pA dampened them. (iv) dCaAP (in blue) precedes the somatic AP (in gray); traces are magnified from the framed APs in (ii) and (iii). (**C**) Somatic current injection, I_{soma} (i), somatic AP trains (ii), and bAP (iii) for similar ranges of current intensity as those shown in (i) of (B). (iv) Somatic AP (in gray) precedes the dendritic bAP (in green); traces are magnified from the framed APs in (ii) and (iii). (**D**) Increase in I_{dend} (i) dampened the dCaAPs' amplitude (ii); vertical tick on each trace marks 50 ms after I_{dend} onset. stim., stimulation; norm., normalized. (**E**) Amplitude of the first dCaAP in each trace against I_{dend} normalized by rheobase (I_{rhe}) for uncoupled dCaAPs (12 dendrites) and exponential fit (dashed line), with a decay constant (τ_{dCaAP}) of 0.39 (median 0.38) in units of rheobase. (**F**) dCaAP amplitudes as in (E) but not normalized by I_{rhe}. Dots in different colors represent dCaAP amplitudes from different cell (12 dendrites) with exponential fit (dashed lines). (**G**) As in (D) but for somatic APs. I_{soma}, (i) and the resulting somatic APs (ii). (**H**) AP amplitude plotted against the normalized somatic input current strength (I_{soma}/I_{rhe}). The amplitude of the somatic AP was fixed and did not depend on I_{soma} for a range of stimuli strengths as in (ii) of (G) (exponential fit with τ_{AP} = 82, units of somatic I_{rhe}). (**I**) Dendritic and somatic activation functions for dCaAPs (blue curve) and for somatic APs (black curve).

soma of the same cells, AP output increased with the input's strength (Fig. 2C). These results are explained by the unusual active properties of dCaAPs. dCaAPs evoked by the dendritic electrode triggered somatic APs near threshold but were suppressed by further increase in the stimulus intensity (Fig. 2B).

The dendritic activation function (namely, the amplitude of dCaAPs as a function of the intensity of the current injection in the dendrite, I_{dend}) reached its maximal value at the rheobase (i.e., for I_{dend} = I_{rhe} where I_{rhe} is the threshold current for triggering a dCaAP) and decayed for stronger I_{dend} (Fig. 2, D to F; 12 uncoupled dCaAPs). The mean width of the dendritic activation function (defined here as the decay constant of a single exponential fit) was 0.39 (0.38 median; in units of I_{rhe}), which indicates that dCaAPs are sharply tuned (highly selective) to a particular input strength. Additionally, L2/3 dendrites were heterogeneous in their activation function threshold and width (Fig. 2F). In contrast, in a similar range of input intensities, somatic APs (Fig. 2, G to H) showed a typical threshold activation function; once a somatic AP was triggered, its amplitude was virtually independent of the input intensity (Fig. 2H). Unlike other dendritic APs in the mammalian neocortex—namely, NMDA spikes (*18*) and dendritic Ca^{2+} APs in layer

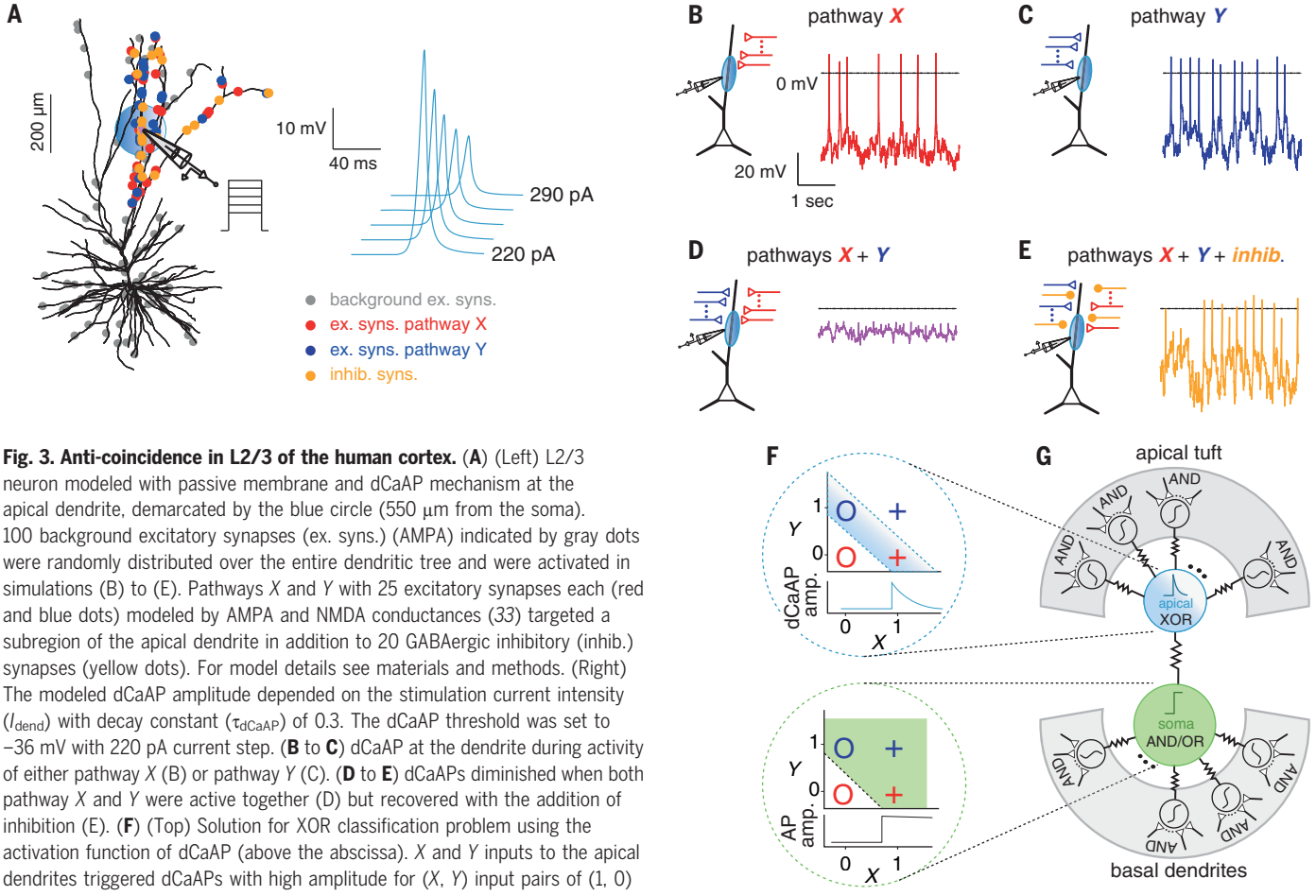


Fig. 3. Anti-coincidence in L2/3 of the human cortex. (**A**) (Left) L2/3 neuron modeled with passive membrane and dCaAP mechanism at the apical dendrite, demarcated by the blue circle (550 µm from the soma). 100 background excitatory synapses (ex. syns.) (AMPA) indicated by gray dots were randomly distributed over the entire dendritic tree and were activated in simulations (B) to (E). Pathways *X* and *Y* with 25 excitatory synapses each (red and blue dots) modeled by AMPA and NMDA conductances (*33*) targeted a subregion of the apical dendrite in addition to 20 GABAergic inhibitory (inhib.) synapses (yellow dots). For model details see materials and methods. (Right) The modeled dCaAP amplitude depended on the stimulation current intensity (I_{dend}) with decay constant (τ_{dCaAP}) of 0.3. The dCaAP threshold was set to −36 mV with 220 pA current step. (**B** to **C**) dCaAP at the dendrite during activity of either pathway *X* (B) or pathway *Y* (C). (**D** to **E**) dCaAPs diminished when both pathway *X* and *Y* were active together (D) but recovered with the addition of inhibition (E). (**F**) (Top) Solution for XOR classification problem using the activation function of dCaAP (above the abscissa). *X* and *Y* inputs to the apical dendrites triggered dCaAPs with high amplitude for (*X*, *Y*) input pairs of (1, 0) and (0, 1), marked by blue circle and red cross, but not for (0, 0) and (1, 1), marked by red circle and blue cross. (Bottom) Solution for OR classification. Somatic AP was triggered for (*X*, *Y*) input pairs of (1, 1), (0, 1), and (1, 0), but not for (0, 0). (**G**) Schematic model of a L2/3 pyramidal neuron with somatic compartment (green) presented as logical AND/OR gate with activation function of somatic AP, apical dendrite compartment as logical XOR gate, and basal and tuft dendritic braches, in gray background, as logical gate AND due to the NMDA spikes (*33*).

5 pyramidal neurons (*15*, *19*–*22*)—that were previously shown to increase with the stimulus strength, the activation function of dCaAPs in L2/3 neurons was sharply tuned to a specific input strength (Fig. 2I).

We used a compartmental model of a L2/3 pyramidal neuron that replicated the phenomenology of the dCaAP behavior in the dendrite to investigate the functional outcome of the dCaAP activation function (for a biophysical model of dCaAPs, see fig. S12). L2/3 pyramidal neuron morphology was digitally reconstructed and modeled in the NEURON (*23*) simulation environment (Fig. 3A). The dCaAP's threshold, width, and amplitude as a function of the input strength were simulated by the sum of current sources with a sigmoidal shape (for details, see materials and methods and Fig. 3A, right panel). To simulate two distinct classes of inputs, pathways *X* and *Y*, we used 25 excitatory synapses for each pathway (Fig. 3A), targeting a subregion of the apical dendrite (blue and red dots in Fig. 3A). Each of these pathways was able to trigger dCaAPs by itself (Fig. 3, B and C). Because of the activation function of the dCaAPs in our simulation, coincident activation of two synaptic input pathways diminished the dCaAP amplitude (Fig. 3D) in contrast to other dendritic APs that amplify coincident dendritic inputs (*24*) [e.g., in layer 5 pyramidal neurons in the rodent neocortex (*25*) or in CA1 neurons of the rodent hippocampus (*26*)]. Our simulation is therefore a simple and explicit demonstration of how the dendritic mechanism observed in human L2/3 pyramidal neurons computes an anticoincident function for multiple input pathways, limiting the number and/or the strength of inputs integrated in the dendrite (for impact on the cell body, see fig. S9). Inhibition (*27*, *28*) placed at the same dendritic subregion (20 GABAergic synapses), in addition to the two excitatory pathways, repolarized the membrane and recovered the amplitude of the dCaAPs [Fig. 3E; (*29*)]. These results suggest that the precise balance between excitation and inhibition is essential for the generation of dCaAPs and indicate a counterintuitive role for inhibition in enhancing the excitability of the dendrite (see also fig. S9, C and D).

It has long been assumed that the summation of excitatory synaptic inputs at the dendrite and the output at the axon can only instantiate logical operations such as AND and OR (*30*). Traditionally, the XOR operation has been thought to require a network solution (*31*, *32*). We found that the dCaAPs' activation function allowed them to effectively compute the XOR operation in the dendrite by suppressing the amplitude of the dCaAP when the input is above the optimal strength (Fig. 2). Thus, on the basis of our results and those of previous studies (*30*, *33*), we consider a model that portrays the somatic and dendritic compartments of L2/3 neurons as a network of coupled logical operators and corresponding activation functions (Fig. 3, F and G). In this model, the XOR operation is performed in the dendrites with dCaAPs, whereas AND/OR operations are performed at the soma and at tuft and basal dendrites

with sodium and NMDA spikes, respectively (*20*, *25*, *34*, *35*). Our findings provide insights into the physiological building blocks that constitute the algorithms of cellular function, which ultimately give rise to the cortical network behavior.

REFERENCES AND NOTES

1. J. DeFelipe, L. Alonso-Nanclares, J. I. Arellano, *J. Neurocytol.* **31**, 299–316 (2002).
2. H. Mohan *et al*., *Cereb. Cortex* **25**, 4839–4853 (2015).
3. Y. Deitcher *et al*., *Cereb. Cortex* **27**, 5398–5414 (2017).
4. J. Waters, M. Larkum, B. Sakmann, F. Helmchen, *J. Neurosci.* **23**, 8558–8567 (2003).
5. M. E. Larkum, J. Waters, B. Sakmann, F. Helmchen, *J. Neurosci.* **27**, 8999–9008 (2007).
6. L. Beaulieu-Laroche *et al*., *Cell* **175**, 643–651.e14 (2018).
7. S. L. Smith, I. T. Smith, T. Branco, M. Häusser, *Nature* **503**, 115–120 (2013).
8. J. J. Moore *et al*., *Science* **355**, eaaj1497 (2017).
9. B. Sivyer, S. R. Williams, *Nat. Neurosci.* **16**, 1848–1856 (2013).
10. S. Crochet, P. Fuentealba, I. Timofeev, M. Steriade, *Cereb. Cortex* **14**, 1110–1121 (2004).
11. Y. Kim, C.-L. Hsu, M. S. Cembrowski, B. D. Mensh, N. Spruston, *eLife* **4**, e06414 (2015).
12. M. E. Larkum, J. J. Zhu, B. Sakmann, *J. Physiol.* **533**, 447–466 (2001).
13. N. L. Golding, N. Spruston, *Neuron* **21**, 1189–1200 (1998).
14. L. M. Palmer *et al*., *Nat. Neurosci.* **17**, 383–390 (2014).
15. M. E. Larkum, J. J. Zhu, B. Sakmann, *Nature* **398**, 338–341 (1999).
16. N. L. Golding, H. Y. Jung, T. Mickus, N. Spruston, *J. Neurosci.* **19**, 8789–8798 (1999).
17. C. Labarrera *et al*., *Cell Rep.* **23**, 1034–1044 (2018).
18. J. Schiller, G. Major, H. J. Koester, Y. Schiller, *Nature* **404**, 285–289 (2000).
19. M. E. Larkum, K. M. M. Kaiser, B. Sakmann, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14600–14604 (1999).
20. M. E. Larkum, T. Nevian, M. Sandler, A. Polsky, J. Schiller, *Science* **325**, 756–760 (2009).
21. J. J. Zhu, *J. Physiol.* **526**, 571–587 (2000).
22. J. Schiller, Y. Schiller, G. Stuart, B. Sakmann, *J. Physiol.* **505**, 605–616 (1997).
23. M. L. Hines, N. T. Carnevale, *Neural Comput.* **9**, 1179–1209 (1997).
24. M. London, M. Häusser, *Annu. Rev. Neurosci.* **28**, 503–532 (2005).
25. A. Polsky, B. W. Mel, J. Schiller, *Nat. Neurosci.* **7**, 621–627 (2004).
26. A. Losonczy, J. C. Magee, *Neuron* **50**, 291–307 (2006).
27. E. Boldog *et al*., *Nat. Neurosci.* **21**, 1185–1195 (2018).
28. A. Gidon, I. Segev, *Neuron* **75**, 330–341 (2012).
29. A. T. Gulledge, G. J. Stuart, *Neuron* **37**, 299–309 (2003).
30. G. M. Shepherd, R. K. Brayton, *Neuroscience* **21**, 151–165 (1987).
31. M. Minsky, S. A. Papert, *Perceptrons: An Introduction to Computational Geometry* (MIT Press, 1969).
32. F. Rosenblatt, *The Perceptron, a Perceiving and Recognizing Automaton: (Project Para)* (Cornell Aeronautical Laboratory, 1957).
33. G. Eyal *et al*., *Front. Cell. Neurosci.* **12**, 181 (2018).
34. P. Poirazi, T. Brannon, B. W. Mel, *Neuron* **37**, 989–999 (2003).
35. G. Testa-Silva, S. Honnuraiah, C. French, J. King, K. Drummond, L. M. Palmer, G. J. Stuart, "NMDA spikes in human neocortex," program no. 463.12, *2018 Neuroscience Meeting Planner*, Society for Neuroscience, San Diego, CA, 3 to 7 November 2018.
36. A. Gidon *et al*., Dendritic action potentials and computation in human layer 2/3 cortical neurons. Zenodo (2020); https://doi.org/10.5281/zenodo.3530043.

ACKNOWLEDGMENTS

We thank U. Schneider for providing the human tissue; L. Kraus, A. Ragot, O. Kruchik, and I. Wolter for assisting with human tissue processing; and S. Grosser and F. J. Barreda Tomás for assisting with confocal imaging. **Funding:** This work was supported by Deutsche Forschungsgemeinschaft DFG: 2112280105 (to T.A.Z.), EXC 257 (to I.V., P.P., and M.E.L.), FOR 2143 (to I.V.), EXC 2049 (to P.F.), LA 3442/3-1 (to M.E.L.), SPP1665 (to M.E.L.), and SFB1078 B2 (to M.E.L.); 7FP Health-F2-602531-2013 DESIRE (to M.H.); Hellenic Foundation for Research and Innovation HFRI and the General Secretariat for Research and Technology GSRT 1357 (to A.P.); Humboldt Foundation Friedrich Wilhelm Bessel Research Award (to P.P.); H2020 European Research Council ERC STG 311435 (to P.P.); H2020 Research and Innovation Programme 720270/HBP SGA1, 785907/HBP SGA2, and 670118/ERC ActiveCortex (to M.E.L.); and EMBO ALTF 1193–2015 (to A.G.).
Author contributions: M.E.L. and A.G. conceptualized and performed the experiments and analysis and wrote the original draft. T.A.Z. performed the 2-p experiments. P.F. and M.H. managed human tissue ethical aspects, delivery, and quality optimization. A.G., A.P., and P.P. conceptualized and created the models. F.B. and I.V. performed visualization, imaging, and morphological reconstruction of recorded neurons. All the authors participated in writing, reviewing, and editing the manuscript.
Competing interests: The authors declare no competing interests.
Data and materials availability: NEURON simulation files are available at https://modeldb.yale.edu/254217 for Fig. 3 and fig. S9 and at https://modeldb.yale.edu/260178 for fig. S12. Reconstructed neurons and all summary graphs with corresponding data are stored at Zenodo (*36*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6473/83/suppl/DC1
Materials and Methods
Figs. S1 to S12
Tables S1 and S2
References (*37*–*51*)

View/request a protocol for this paper from *Bio-protocol*.

30 April 2019; accepted 22 November 2019
10.1126/science.aax6239

ARCHAEOLOGY

Cooked starchy rhizomes in Africa 170 thousand years ago

Lyn Wadley[1]*, Lucinda Backwell[1,2], Francesco d'Errico[3,4], Christine Sievers[5]

Plant carbohydrates were undoubtedly consumed in antiquity, yet starchy geophytes were seldom preserved archaeologically. We report evidence for geophyte exploitation by early humans from at least 170,000 years ago. Charred rhizomes from Border Cave, South Africa, were identified to the genus *Hypoxis* L. by comparing the morphology and anatomy of ancient and modern rhizomes. *Hypoxis angustifolia* Lam., the likely taxon, proliferates in relatively well-watered areas of sub-Saharan Africa and in Yemen, Arabia. In those areas and possibly farther north during moist periods, *Hypoxis* rhizomes would have provided reliable and familiar carbohydrate sources for mobile groups.

Ancient hunting strategies receive more attention from archaeologists than plant collecting, because plant preservation is often poor in archaeological sites. A plant diet, though sometimes invisible, must have contributed substantially to food security in the past, as reported for hunter-gatherers in Africa during the last century (*1*). African venison is especially lean in the dry season; thus, human populations able to supplement meat diets with carbohydrate or fat avoided malnourishment (*2*, *3*). Geophytes (corms, bulbs, tubers, and rhizomes) store starch in their underground organs, and these underground portions become sources of carbohydrate for humans and those animals able to excavate them. Modern collecting of edible geophytes in South Africa demonstrates that a gatherer's daily caloric requirement can be met within 2 hours (*4*), and such foods may become dietary staples. Cooking increases digestibility of meat and plant food, reduces toxicity, and, in the case of geophytes, has a considerable softening effect (*5*), eases peeling and chewing, and enhances glucose availability (*6*, *7*).

Before the use of fire, hominins may have eaten geophytes raw, especially Cyperaceae and aquatic plants, though some of these first required pulverizing (*8*). No geophyte remains have been recovered from early, pre-fire sites. Instead, circumstantial evidence is from sources such as isotopes, DNA, or dental calculus (*9*–*14*). Seeds from several edible geophytes and aquatic plants were found in samples from 780 thousand years (ky) ago at Gesher Benot Ya'aqov in Israel (*15*); the samples contained seeds rather than the geophytes themselves, perhaps because no trace remains after consumption of geophytes. At Klasies River, South Africa, 120-ky-old charred parenchyma fragments from unidentified geophytes imply cooking (*16*). Holocene layers of Bushman Rock Shelter and Melkhoutboom, South Africa, yielded geophytes including *Hypoxis* spp. (*17*, *18*). Here, we present earlier evidence, from at least 170 ky ago, for the cooking of identified edible rhizomes from Border Cave, northern KwaZulu-Natal, South Africa (*19*) (fig. S1).

Border Cave formed in a Lebombo Mountain cliff facing west across eSwatini (Fig. 1A). The cave was extensively excavated from early in the 1900s (*20*), but little attention was given to its botanical remains, apart from preliminary studies of 40-ky-old seeds and leaves (*21*) and chemical identifications of plant poison and resin (*22*). The cave has alternating brown sand (BS) and white ash (WA) stratigraphic members (*21*), and rhizomes were recovered from the 5 Brown Sand (5 BS) and 4 White Ash (4 WA) samples. Discrete layers are discernible in each member (Fig. 1B). The majority of identified 4 WA rhizomes come from White 8 to 5 (Table 1), which are combustion features near the base of the member. Only humans could have transported whole rhizomes from the field to the cave. The Border Cave specimens were preserved because they were charred and presumably because they were lost while roasting in the ashes, from which they were recovered archaeologically (Fig. 2). Nothing indicates that the rhizomes were pulverized

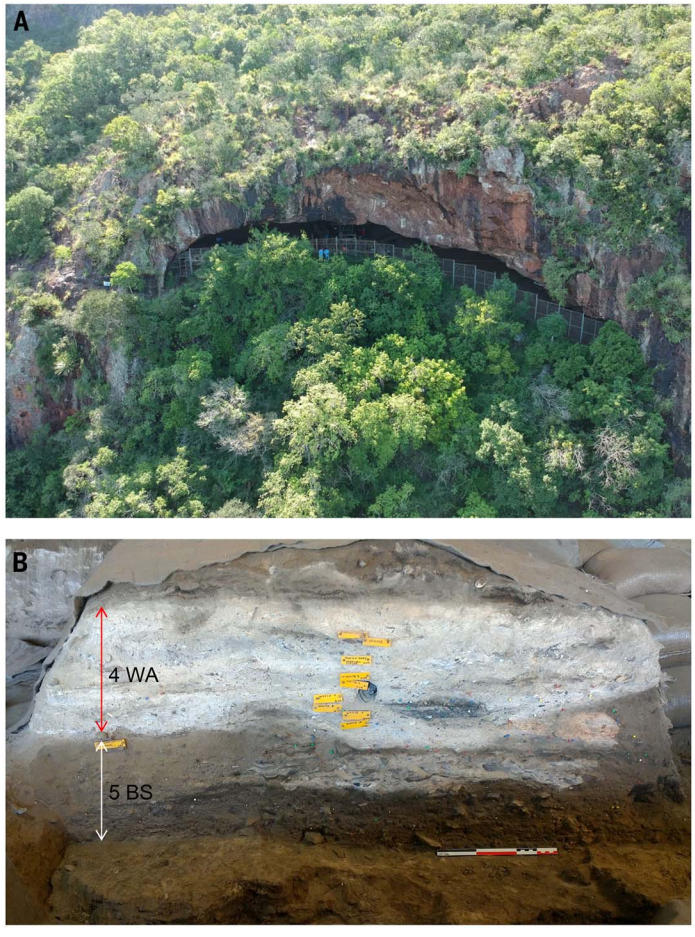


Fig. 1. Border Cave entrance and stratigraphy in Members 5 BS and 4 WA. (**A**) Border Cave perched on a cliff in the Lebombo Mountains. (**B**) Border Cave stratigraphy, excavated from 2015 to 2018, from Members 5 BS and 4 WA in Squares N108 E113 and N109 E113. Note the stacked combustion features in 4 WA. Scale bar, 30 cm.

[1]Evolutionary Studies Institute, University of the Witwatersrand, Johannesburg, South Africa. [2]Instituto Superior de Estudios Sociales, ISES-CONICET, Tucumán, Argentina. [3]Centre National de la Recherche Scientifique, UMR 5199—PACEA, Université de Bordeaux, Bâtiment B2, Allée Geoffroy Saint Hilaire, CS 50023, 33615 Pessac, France. [4]SFF Centre for Early Sapiens Behaviour (SapienCE), University of Bergen, Øysteinsgate 3, Postboks 7805, 5020 Bergen, Norway. [5]School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, Johannesburg, South Africa.
*Corresponding author. Email: lyn.wadley@wits.ac.za

(*23*), but circumstantial evidence for cooking is compelling. The spatial context of the rhizomes in ash rather than adjacent sediment is significant. Further support for cooking comes from amylase gene analysis results, which indicate that a high starch diet, possibly involving processing and/or cooking of carbohydrate-rich geophytes by early humans, was already in place by the Middle Pleistocene (*24*). Cooking enables dietary diversity (*25*), and transporting geophytes to a home base like Border Cave facilitates both food processing and sharing.

The early Middle Stone Age occupations at Border Cave were dated by electron spin resonance on tooth enamel to 161 ± 10 and 144 ± 11 ky ago for 5 BS and 168 ± 5, 115 ± 8, and 113 ± 5 ky ago for 4 WA (means ± standard errors) (*26*). Bayesian analysis to obtain 95.4% probability ranges for boundaries between members indicated that Member 5 BS occupations may have commenced around 177 ky ago, while those in 4 WA began about 150 ky ago and terminated no later than 99 ky ago (*27*).

The term rhizome is used here (supplementary text), following the terminology of Wiland-Szymańska and Adamski (*28*) and Singh (*29*). The Border Cave 5 BS and 4 WA rhizomes comprise 55 complete, charred specimens (Table 1) that seem to belong to the same taxon, as well as fragments of charred parenchyma that cannot be securely identified (table S1). They are typically globose with a convex or slightly conical base, a depressed center on the proximal surface, radial splits in the ground tissue, rings from leaf scars or fiber on the outer circumference (Fig. 2), and root traces emerging from fiber sheaths within the cortex (Fig. 3B). High magnification using scanning electron microscopy (SEM) reveals distinctive anatomical features, notwithstanding the deterioration caused by charring and mechanical abrasion by cave sediment (supplementary text). Vascular bundles mostly comprise elongated clusters of xylem vessels (12 to >20) with thickened walls of scalariform tissue (Fig. 3D). Parenchyma cells, when not fused into a solid carbon mass, have geometric or oval shapes. The outer ground tissue (cortex) contains many bundles of needle-like raphides of calcium oxalate (Fig. 3F) that are also present in lower frequencies in the inner ground tissue (pith). All of these morphological and anatomical features match those of modern *Hypoxis* L. spp. rhizomes (Fig. 3, A, C, and E; figs. S2 and S3). Attributes of geophytes from other families appropriate to the area did not match the Border Cave specimens (supplementary text; table S2).

The morphology of the Border Cave rhizomes is suggestive of *Hypoxis angustifolia* Lam. (Fig. 4, A, B, and C) rather than the tiny *H. filiformis* Baker or large *H. hemerocallidea* Fisch., C.A.Mey. & Avé-Lall. The relatively small, slender-leafed *Hypoxis* taxa, such as

Table 1. Border Cave whole rhizome frequencies in Members 5 BS and 4 WA.

Member	Layer	Frequency of whole rhizomes
4 WA	Top and Pinkish Grey	6
	White 1, 2, and 3	8
	White 5 and 6	14
	White 7 and 8	13
	Reddish Brown lan, Dark Brown pit	3
5 BS	Very Dark Grey and Slump	6
	Dark Greyish Brown	5
Total		55

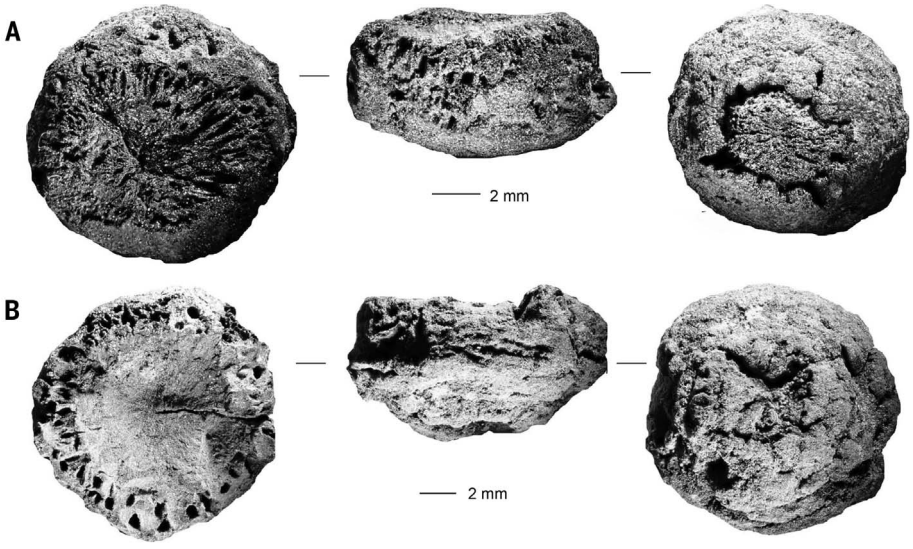


Fig. 2. Two whole Border Cave charred rhizomes. (**A**) Three views of Border Cave rhizome BC 17 from Square N108 E114, Member 4 WA, layer White 6.1 Idaho (#9577). The proximal end (top left) has been abraded, exposing radial splits caused by charring the fresh organ. Scale bar, 2 mm. (**B**) Three views of Border Cave rhizome BC 6 from Square N108 E114, Member 4 WA, layer White 6 Idaho (#9433). The rhizome proximal end (left) has a marked central depression and cavities, some of which are root cavities on the circumference. The rhizome profile (center) has a few raised ring scars, and the distal end (right) has a wrinkled, convex base. Scale bar, 2 mm.

H. angustifolia, *H. argentea* Harv. ex Baker, and *H. filiformis*, have rhizomes with white (Fig. 4) or pale yellow flesh. They are more palatable than the orange-fleshed rhizomes from plants such as *H. hemerocallidea* (*29*) and are therefore favored as food. While they are edible raw, *Hypoxis* rhizomes have high fracture toughness until they are cooked (*5*). The rhizomes are nutritious with an energy value of approximately 500 KJ/100 g, supplemented by essential vitamins and minerals (*17*). *H. hemerocallidea* rhizomes have constant carbohydrate composition year-round, though soluble sugars are slightly reduced during the resting season (*30*).

H. angustifolia, *H. argentea*, *H. filiformis*, and *H. hemerocallidea* grow in the Border Cave area, but recent overexploitation has reduced their populations. *H. angustifolia* is gregarious, propagates readily from rhizome side shoots (Fig. 4C), and retains visibility year-round, unlike deciduous *Hypoxis* species. It thrives in a variety of modern habitats and is thus likely to have had a wide distribution in the past as it does today. It occurs in sub-Saharan Africa, Sudan (to about 13°N), some Indian Ocean islands, and as far afield as Yemen (Fig. 4D) (*28*, *31*). Its modern occurrence in Yemen may indicate wider distribution of the rhizome during previous periods of humid conditions.

The Border Cave discovery is early evidence of cooked starchy plant food. The wide distribution of *Hypoxis*, particularly the small, palatable *Hypoxis angustifolia* rhizome that

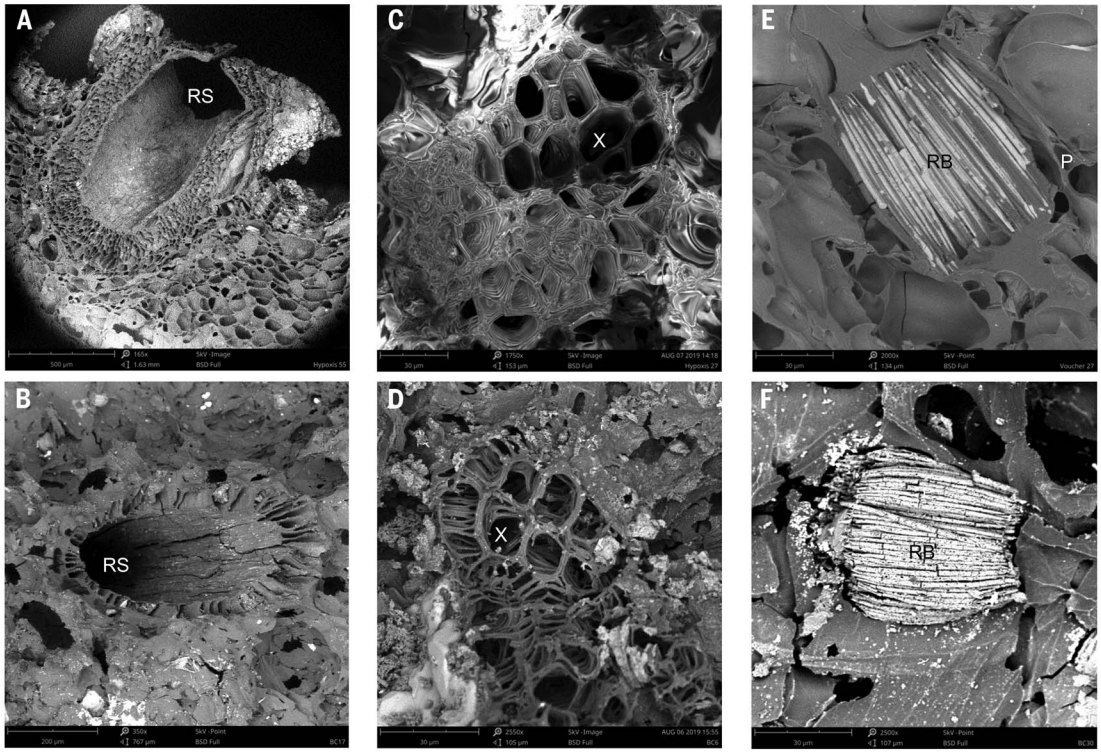


Fig. 3. Selected anatomical features of modern charred *Hypoxis* rhizomes compared with Border Cave rhizomes. (**A**) Modern charred *Hypoxis* rhizome #55 with rhizodermal sheath. Transverse section (TS) scale, 500 μm. (**B**) Border Cave rhizome BC 17 with rhizodermal sheath. TS scale, 200 μm. (**C**) Modern charred *Hypoxis* rhizome #27, cluster of xylem vessels with walls of scalariform tissue. TS scale, 30 μm. (**D**) Border Cave BC 6, cluster of xylem vessels with walls of scalariform tissue. TS scale, 30 μm. (**E**) Modern, charred *Hypoxis* rhizome #27 with raphide bundle. TS scale, 30 μm. (**F**) Border Cave rhizome BC 30, raphide bundle. TS scale, 30 μm. X, xylem; P, parenchyma; RB, raphide bundle; RS, rhizodermal sheath.

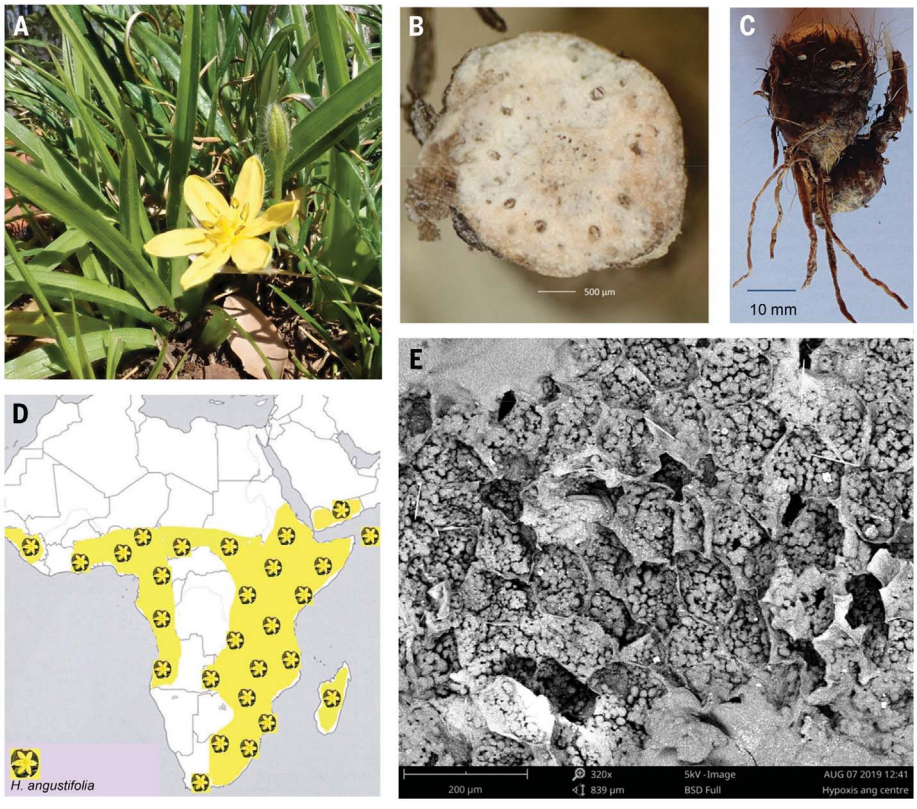


Fig. 4. Modern *Hypoxis angustifolia* field appearance, morphology, anatomy, and distribution. (**A**) *H. angustifolia* habit. (**B**) *H. angustifolia* fresh rhizome, cut transversely to reveal white flesh. Scale bar, 500 μm. (**C**) *H. angustifolia* mature fresh rhizome (left) with developing offshoot (right). Scale bar, 10 mm. (**D**) Simplified distribution of *H. angustifolia* through Africa, Yemen, and Indian Ocean islands. (**E**) *H. angustifolia* fresh rhizome cut transversely to expose parenchyma cells filled with starch grains. Scale bar, 200 μm.

grows gregariously in many habitats, implies that it could have provided a reliable, familiar staple food source for early humans moving within or out of Africa. Monocotyledons also provided food security when people moved through southern Europe during the Palaeolithic (*32*). Certain raw materials found in the Middle Stone Age site of Olorgesaile, East Africa, had distant origins by ≥295 to ~320 ky ago, implying that mobile networks were long-standing (*33*). Border Cave is remote from dispersal routes out of Africa, yet the site contributes data (*19*) for assessing the ease with which early *Homo sapiens* could travel within the continent.

REFERENCES AND NOTES

1. G. B. Silberbauer, *Hunter and Habitat in the Central Kalahari Desert* (Cambridge Univ. Press, 1981).
2. P. J. Butterworth, P. R. Ellis, M. Wollstonecroft, in *Wild Harvest: Plants in the Hominin and Pre-agrarian Human Worlds*, K. Hardy, Kubiak-Martens, Eds. (Oxbow, Oxford, 2016), pp. 31–54.
3. J. D. Speth, K. A. Spielmann, *J. Anthropol. Archaeol.* **2**, 1–31 (1983).
4. E. Singels *et al.*, *J. Hum. Evol.* **101**, 79–89 (2016).
5. N. J. Dominy, E. R. Vogel, J. D. Yeakel, P. Constantino, P. W. Lucas, *Evol. Biol.* **35**, 159–175 (2008).
6. R. Wrangham, *Curr. Anthropol.* **58** (S16), S303–S313 (2017).
7. L. Wandsnider, *J. Anthropol. Archaeol.* **16**, 1–48 (1997).
8. M. M. Wollstonecroft, P. R. Ellis, G. C. Hillman, D. Q. Fuller, *Veg. Hist. Archaeobot.* **17**, 19–27 (2008).
9. M. Sponheimer, J. A. Lee-Thorp, *Science* **283**, 368–370 (1999).
10. C. R. Peters, J. C. Vogel, *J. Hum. Evol.* **48**, 219–236 (2005).
11. J. Lee-Thorp *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 20369–20372 (2012).
12. K. Hardy, J. Brand-Miller, K. D. Brown, M. G. Thomas, L. Copeland, *Q. Rev. Biol.* **90**, 251–268 (2015).
13. A. G. Henry, A. S. Brooks, D. R. Piperno, *J. Hum. Evol.* **69**, 44–54 (2014).
14. L. S. Weyrich *et al.*, *Nature* **544**, 357–361 (2017).
15. Y. Melamed, M. E. Kislev, E. Geffen, S. Lev-Yadun, N. Goren-Inbar, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14674–14679 (2016).
16. C. Larbey, S. M. Mentzer, B. Ligouis, S. Wurz, M. K. Jones, *J. Hum. Evol.* **131**, 210–227 (2019).
17. L. Wadley, *Hunters and Gatherers of the Later Stone Age, Southern Transvaal* (Cambridge Monographs in African Archaeology 25, Oxford, 1987).
18. H. J. Deacon, *Where Hunters Gathered* (South African Archaeological Society Monograph Series 1, 1976).
19. Materials and methods are available as supplementary materials.
20. L. R. Backwell *et al.*, *J. Field Archaeol.* **43**, 417–436 (2018).
21. P. B. Beaumont, thesis, University of Cape Town, South Africa (1978).
22. F. d'Errico *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 13214–13219 (2012).
23. L. González Carretero, M. Wollstonecroft, D. Q. Fuller, *Veg. Hist. Archaeobot.* **26**, 415–432 (2017).
24. C. E. Inchley *et al.*, *Sci. Rep.* **6**, 37198 (2016).
25. G. Hillman, M. Wollstonecroft, in *Archaeology of African Plant Use*, C. J. Stephens, S. Nixon, M. A. Murray, D. Q. Fuller, Eds. (Left Coast Press, Walnut Creek, 2014), pp. 37–49.
26. R. Grün, P. Beaumont, P. V. Tobias, S. Eggins, *J. Hum. Evol.* **45**, 155–167 (2003).
27. A. R. Millard, *Quat. Geochronol.* **1**, 159–166 (2006).
28. J. Wiland-Szymańska, Z. Adamski, J. Wiland-Szymanska, *Novon* **12**, 142–151 (2002).
29. Y. Singh, thesis, University of Pretoria, South Africa (2000).
30. J. W. Bews, J. E. Vanderplank, *Ann. Bot.* **os-44**, 689–719 (1930).
31. J. Wiland-Szymańska, *Biodivers. Res. Conserv.* **14**, 1–129 (2009).
32. M. Jones, in *The Evolution of Hominin Diets: Integrating Approaches to the Study of Palaeolithic Subsistence*, J.-J. Hublin, M. P. Richards, Eds. (Springer, 2009), pp. 171–180.
33. A. S. Brooks *et al.*, *Science* **360**, 90–94 (2018).

ACKNOWLEDGMENTS

We thank M. Bamford, N. Crouch, S. Simm, H. Glen, R. Glen, and L. Kubiak-Martens for comments and botanical assistance and two anonymous reviewers for useful suggestions. Border Cave was excavated with permit #SAH 15/7645 (to L.B.) from Amafa. The comparative plant collection was made with permit #OP4367/2017 (to C.S.) from Ezemvelo KZN Wildlife. G. Pettit supplied some geophytes. The Evolutionary Studies Institute provided resources. A. Kruger produced the drone image. A. Millard provided data from his 2006 Bayesian analysis of ESR dates. **Funding:** L.B. received funding from the National Geographic Society (NGS-54810R-19) and Centre of Excellence (OP2015/04LB). The Research Council of Norway Centres of Excellence, SFF Centre for Early Sapiens Behavior (SapienCE) (262618) partly funded F.D. **Author contributions:** The excavation was conceived by L.B. and F.D. and largely funded by L.B. Botanical investigation, formal analysis, and project conceptualization were undertaken by L.W. and C.S. Writing was by L.W. with contributions from C.S., L.B., and F.D. All authors reviewed and edited the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data and materials are included in the text and supplementary materials. The Border Cave excavated material (by L.B.) is housed in the Evolutionary Studies Institute, University of the Witwatersrand, Johannesburg, South Africa (Border Cave 2015/2019).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6473/87/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S3
Tables S1 and S2
References (*34*–*59*)

View/request a protocol for this paper from *Bio-protocol*.

22 September 2019; accepted 5 November 2019
10.1126/science.aaz5926

PLANT SCIENCE

Evolution of carnivorous traps from planar leaves through simple shifts in gene expression

Christopher D. Whitewoods[1]*, Beatriz Gonçalves[1]*, Jie Cheng[1,2,3]*, Minlong Cui[4], Richard Kennaway[1], Karen Lee[1], Claire Bushell[1], Man Yu[1], Chunlan Piao[4], Enrico Coen[1]†

Leaves vary from planar sheets and needle-like structures to elaborate cup-shaped traps. Here, we show that in the carnivorous plant *Utricularia gibba*, the upper leaf (adaxial) domain is restricted to a small region of the primordium that gives rise to the trap's inner layer. This restriction is necessary for trap formation, because ectopic adaxial activity at early stages gives radialized leaves and no traps. We present a model that accounts for the formation of both planar and nonplanar leaves through adaxial-abaxial domains of gene activity establishing a polarity field that orients growth. In combination with an orthogonal proximodistal polarity field, this system can generate diverse leaf forms and account for the multiple evolutionary origins of cup-shaped leaves through simple shifts in gene expression.

Leaves come in many shapes and sizes. Most consist of planar sheets of cells that harvest light for photosynthesis. Formation of these leaves depends on adaxial and abaxial domains of gene activity in leaf primordia (*1*, *2*). However, the mechanism by which these domains generate sheet-like development is unclear. It is unknown whether growth is oriented by the adaxial-abaxial (ad-ab) boundary throughout the leaf or solely at the epidermis. It is also unclear how orientations of growth and cell division are specified and whether growth orients the plane of division or the plane of division orients growth. Finally, it is unclear how the system for planar leaf development has been modified to generate nonplanar leaves, such as filiform (needle-like) leaves and cup-shaped leaves of carnivorous plants, which have evolved multiple times independently (*3*).

Computational models for formation of flat or cup-shaped leaves have been proposed based on cell divisions being induced by the epidermal ad-ab boundary, with the plane of division orienting growth (*4*, *5*). However, these models are not easily reconciled with observations that cell divisions occur throughout the leaf lamina, not solely at the margin (*6*–*8*). Here, we suggest an alternative mechanism based on the analysis of ad-ab genes in trap and filiform leaf development of the humped bladderwort, *Utricularia gibba* (Lentibulariaceae), an aquatic carnivorous plant (Fig. 1, A to F).

Each *U. gibba* leaf consists of several filiform leaflets (Fig. 1, B, E and F) and may bear a trap (Fig. 1, D and F). At early developmental stages, organ primordia are dome-shaped (Fig. 1G). On the basis of morphology alone, it is unclear at this stage whether these primordia will become leaflets or traps. At later stages, leaflet primordia form tapering cylinders that grow to be slightly wider than thick and curve longitudinally toward the apex (Fig. 1, H to J). By contrast, trap primordia are curved in both longitudinal and transverse sections and consist of three cell layers (Fig. 1K). The inner layer is positioned adaxially (facing the spiral apex, the right-hand side of the images in Fig. 1). As development progresses, the trap grows to a near-spherical shape with a closed mouth (Fig. 1, L to M, white arrowheads), and a two-cell-thick trap door grows out near the dorsal lip (Fig. 1, N and O, orange arrowheads). Over a 20-fold increase in trap length (about 400-fold increase in area), lamina thickness only doubles (Fig. 1, K to O, and fig. S1), resulting in a curved sheet.

To define the ad-ab domains in *U. gibba* (*9*, *10*), we identified homologs (named with Ug prefix) of the adaxially expressed *PHV* and *PHB* genes and abaxially expressed *FIL* and *KAN* genes (*1*, *11*). Before trap and leaflet primordia morphologies clearly diverged, *UgPHV1* was expressed on the adaxial side (black arrowhead in Fig. 2A) and was more restricted in some primordia (yellow arrowhead in Fig. 2A). In leaflet primordia at later stages, *UgPHV1* and *UgFIL1* were expressed on the adaxial and abaxial sides, respectively (Fig. 2, B to E, and fig. S3, A to J). In trap primordia, *UgPHV1* was expressed in the innermost cell layer, extending into the inner side of the trap door (Fig. 2, F to I, and fig. S3, K to P). *UgFIL1* and *UgKAN1* were expressed in the outer layers (Fig. 2, J to Q, and fig. S3, Q to V), although only *UgKAN1* was expressed in the outer ventral region (yellow arrowheads in Fig. 2, N to Q). Thus, the adaxial and abaxial domains of a planar leaf broadly correspond to the inner and outer regions of the trap, respectively. Similar findings were reported for the *Sarracenia purpurea* trap, although *KAN* expression was not detected (*5*).

To determine whether the observed expression patterns have functional importance, we induced ectopic expression of microRNA-resistant *UgPHV1* under the control of the 35S promoter, using a Cre-Lox system (*HS-UgPHV1* plants, see methods for details). After extended heat shock, green fluorescent protein fluorescence and in situ hybridization confirmed ectopic induction throughout the tissue (fig. S4).

To determine how ectopic *UgPHV1* affected development, tissues were imaged daily after induction (Fig. 3, A to D). At 7 days post-induction (Fig. 3D), the main axis could be divided into three regions. (i) An upper region, encompassing the apex and leaves (Fig. 3D, red). The normal spiral organization of the apex had been replaced by an open linear structure (Fig. 3, H to M), containing no trap primordia and only radially symmetrical leaflets (fig. S5). The leaves below the apex bore no traps or bore small malformed traps (Fig. 3D and fig. S6, C to E). Tracing this upper region back through the sequence of daily images showed it derived from primordia located within the spiral apex at the time of induction (Fig. 3C and fig. S6B). (ii) A middle region (Fig. 3D, blue) derived from leaves bearing small traps (80 to 200 μm long) at the time of induction (Fig. 3C). This region had normal leaves bearing traps up to 300 μm long (Fig. 3D, white arrowheads, and fig. S6B), with thick walls and malformed trap doors (fig. S6, F to H). (iii) A lower region (Fig. 3D, black) derived from leaves bearing traps that were longer than 200 μm at the time of induction (Fig. 3C). This region contained normal leaves and traps (Fig. 3D, black arrowheads, and fig. S6B).

Thus, ectopic expression of *UgPHV1* in early primordia leads to loss of trap development (Fig. 3, E and F) and generation of radialized organs, similar to the effect of ectopic *PHV* and *PHB* expression in *Arabidopsis* (*11*–*13*). Ectopic *UgPHV1* expression in later trap primordia, but before traps are 200 μm long, leads to aberrant trap development and growth arrest (Fig. 3G and fig. S6). Ectopic *UgPHV1* expression after this stage has no effect, although this may be due to inefficiency of induction in older traps (fig. S4). These results indicate that restricted *UgPHV1* expression is necessary for initiation and maintenance of trap development.

To explore how domains of ad-ab identity may control leaf morphogenesis, we modeled

[1]Department of Cell and Developmental Biology, John Innes Centre, Norwich Research Park, Colney Lane, Norwich NR4 7UH, UK. [2]State Key Laboratory of Systematic and Evolutionary Botany, CAS Center for Excellence in Molecular Plant Sciences, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China. [3]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China. [4]College of Agriculture and Food Science, Zhejiang Agriculture and Forestry University, Linan 311300, Zhejiang, China.
*These authors contributed equally to this work.
†Corresponding author. Email: enrico.coen@jic.ac.uk

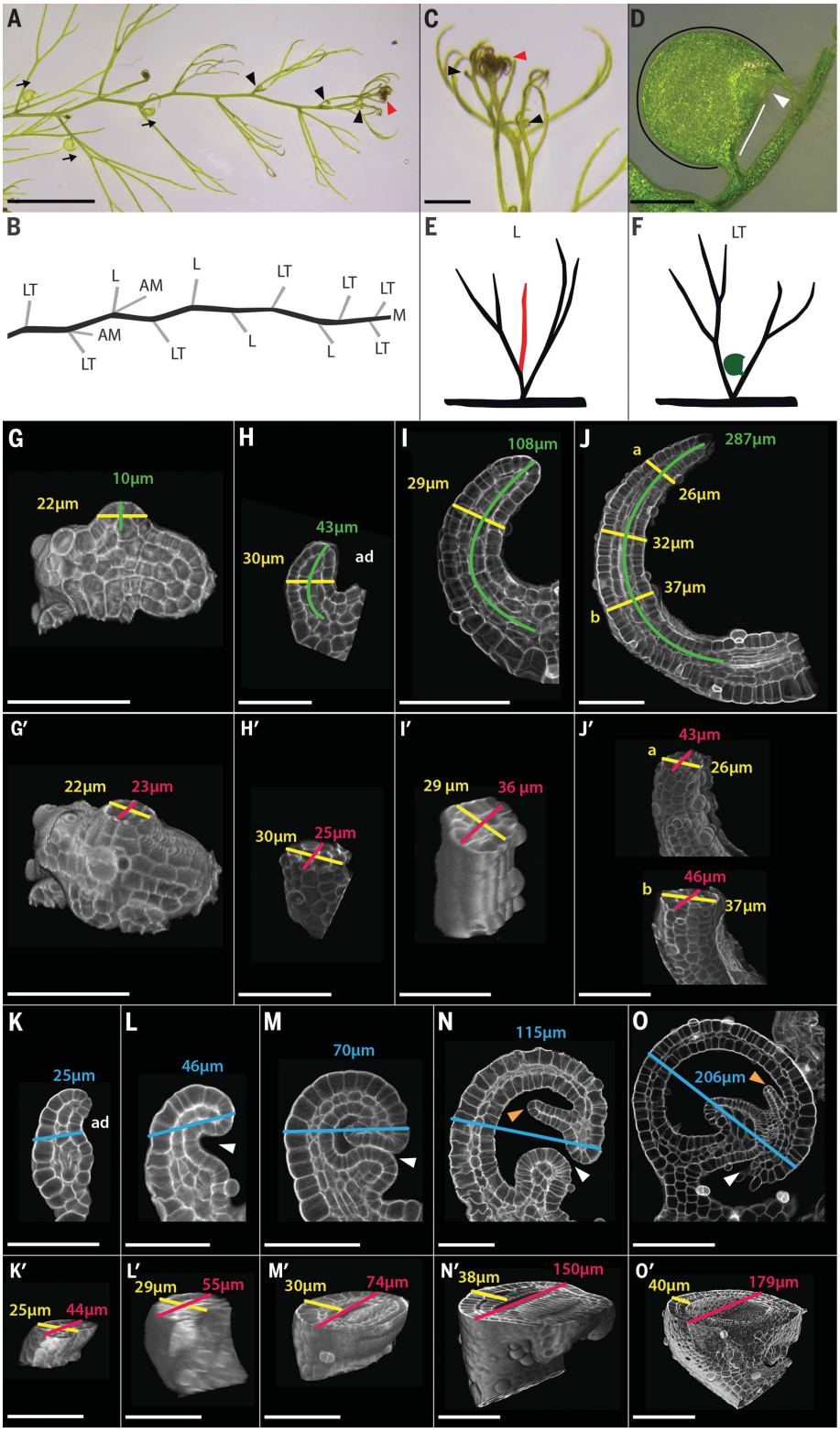


Fig. 1. *U. gibba* morphology. (**A**) Stolon with circinnate apex (red arrowhead) bearing leaves and traps (black arrowheads) oriented adaxially (black arrows). (**B**) Schematic of (A) showing stolons (black) and leaves (gray). L, leaflet-bearing leaf; LT, trap-bearing leaf; AM, axillary meristem; M, apical meristem. (**C**) Magnification of apex shown in (A). (**D**) Trap showing dorsal (black line) and ventral (white line) sides and mouth (white arrowhead). (**E**) Diagram of a leaflet-bearing leaf in (B), with one leaflet in red. (**F**) Diagram of a trap-bearing leaf in (B), with a trap in green. (**G** and **G**) Confocal micrographs of early primordium, in longitudinal (G) and transverse section viewed obliquely (G′). (**H** to **J**) Confocal micrographs of a developmental series of leaflets in longitudinal [(H) to (J)] and transverse sections viewed obliquely [(H′) to (J′)]. (**K** to **O**) Confocal micrographs of a developmental series of traps in longitudinal [(K) to (O)] and transverse sections viewed obliquely [(K′) to (O′)]. In (G) to (O′), green lines mark primordium and leaf length; yellow lines mark primordium, leaf, and trap thickness; red lines mark primordium, leaf, and trap width; and blue lines mark trap length. Measurements are given in the same colors as the lines (see fig. S1 for additional data). In (N) and (O), orange arrowheads mark the trap door, and white arrowheads mark the trap mouth. In (D), (H) to (O′) organs are shown with the adaxial side to the right, marked "ad" in (H) and (K). Scale bars are 5 mm in (A), 1 mm in (C), 500 μm in (D), 50 μm in [(G), (G′), (H), (H′), (I), (I′), (J′), and (K) to (N′)], and 100 μm in [(J), (O), and (O′)]. The image in (N) was previously published in (*26*).

primordia as mechanically connected viscoelastic volumes of material, with growth oriented by a polarity field (*14*). We began with a hemispherical ground state, representing a radialized leaf primordium without a history of ad-ab activity, with polarity pointing proximodistally (red arrows in Fig. 4A). For simplicity, the field was implemented by taking the gradient of a morphogen diffusing from the base toward the tip. In reality, such fields most likely reflect cellular polarity rather than a continuous field (*15*, *16*). Specifying a higher growth rate parallel to, rather than perpendicular to, the polarity generated a tapering cylinder, corresponding to the phenotype

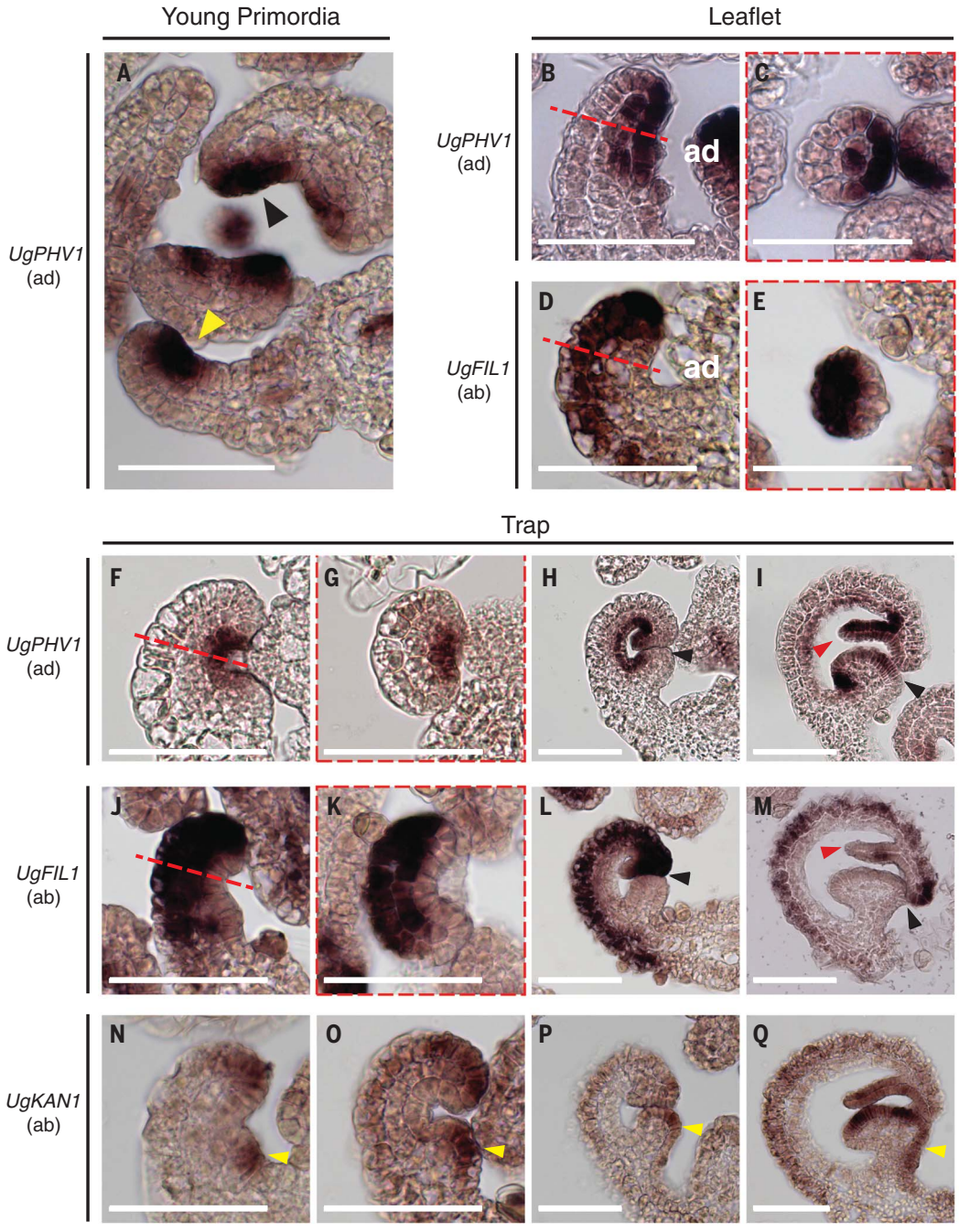


Fig. 2. *UgPHV1* and *UgFIL1* expression patterns in *U. gibba* leaflet and trap development. (**A**) *UgPHV1* expression in young primordia. Black and yellow arrowheads mark primordia that show extended and restricted domains of the adaxial marker, respectively. (**B** to **E**) *UgPHV1* [(B) and (C)] and *UgFIL1* [(D) and (E)] in leaflet primordia. (**F** to **I**) *UgPHV1* during trap development, in longitudinal [(F), (H), and (I)] and cross (G) sections. (**J** to **M**) *UgFIL1* expression during trap development in longitudinal [(J), (L), and (M)] and cross (K) sections. (**N** to **Q**) *UgKAN1* expression during trap development in longitudinal sections. The dashed red lines in (B), (D), (F), and (J) indicate the angle of cross sections shown in (C), (E), (G), and (K), respectively. Trap mouth (black arrowheads), trap door (red arrowheads), and ventral regions (yellow arrowheads) are indicated. In all panels, the adaxial side is shown to the right, marked "ad" in (B) and (D). Scale bars are 50 μm. See an expanded version of this figure in fig. S3.

of radialized mutants (Fig. 4, A to C, and movie S1).

To model formation of a planar leaf, we invoked a second polarity field, orthogonal to the first, termed the orthoplanar polarity field (black arrows in Fig. 4D). This field was implemented by taking the gradient of a morphogen diffusing from the outer surface of the primordium toward the junction between the adaxial and abaxial domains throughout the midplane. Evidence for a midplane domain playing a role in lamina formation comes from analysis of the *WOX-PRS* module (*17*–*20*). Three growth rates could then be specified: parallel to proximodistal polarity (K_{pd}), parallel to the orthoplanar polarity (K_{op}), and perpendicular to both (K_{per}) (*14*) (Fig. 4E).

Setting K_{op} to be much lower than K_{pd} and K_{per} generated a flattened sheet, corresponding to a planar leaf (Fig. 4, F and G, and movie S2). Similar results were obtained using elliptical or elongated initial primordium shapes, illustrating the robustness of the model (fig. S7). Low K_{per} as well as K_{op} generated a tapering elliptic cylinder, corresponding to a leaflet of *U. gibba* (Fig. 4, H and I, and movie S3). Thus, the formation of a planar leaf can be accounted for by having relatively low K_{op}, and the formation of a filiform leaf by both low K_{op} and low K_{per}.

To determine the effect of a more confined adaxial domain, as observed in some *U. gibba* primordia (yellow arrowhead in Fig. 2A), we restricted the domain to a small region on one side of the primordium while keeping the same values of K_{per}, K_{op}, and K_{pd} as in the planar leaf model (Fig. 4, J and L). Running this model generated a cup-shaped sheet

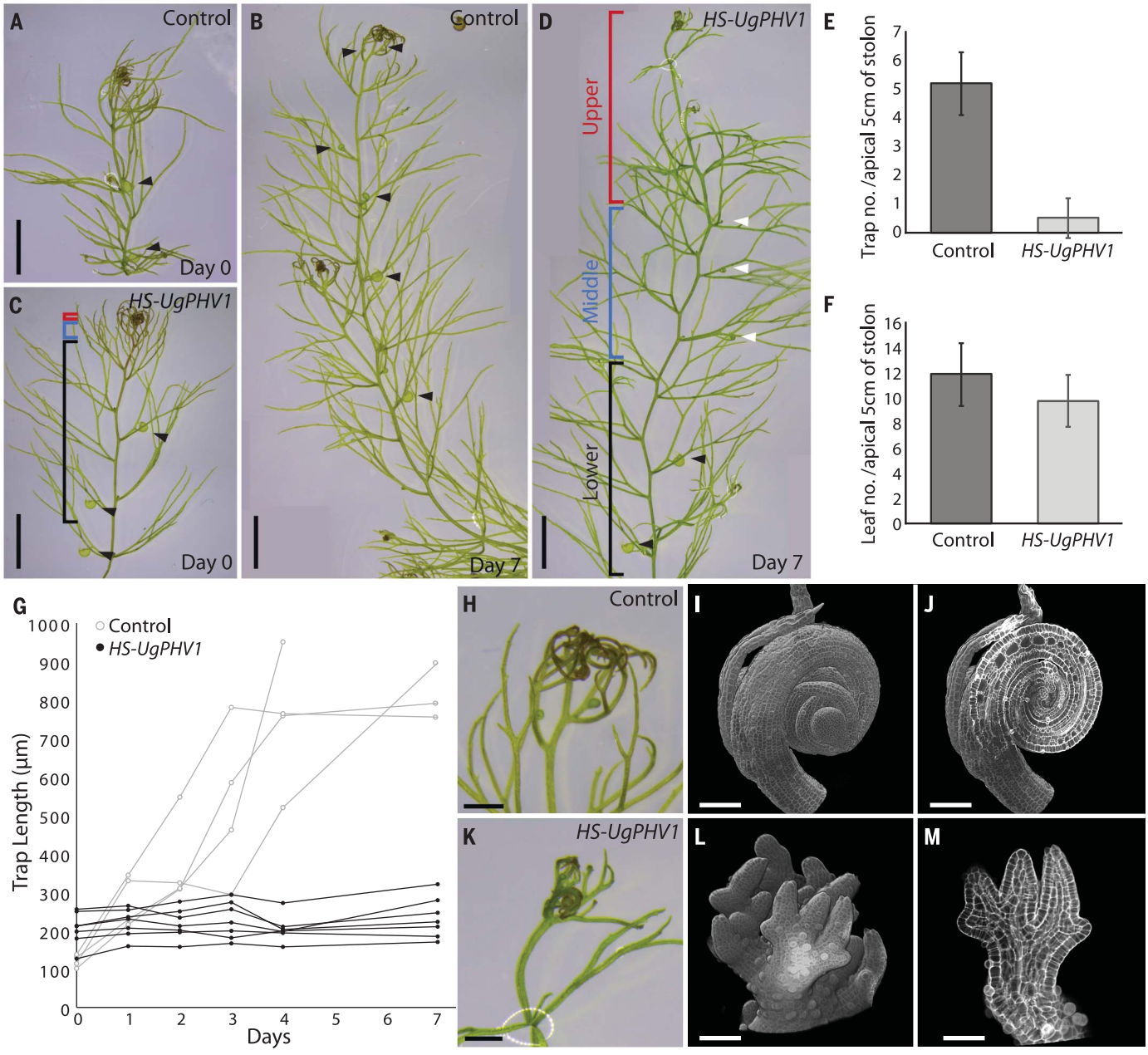


Fig. 3. Induction of *UgPHV1* prevents trap formation. (**A** to **F**) Uninduced plants (control) produce traps (black arrowheads) at regular intervals [(A) and (B)], whereas induction of *UgPHV1* prevented new trap formation [(C), (D), and (E)] and trap growth (white arrowheads) but not leaf formation (F). (**G**) Upon induction, traps smaller than 200 μm did not grow. (**H** to **J**) Apex of control plants showing circinate structure in confocal images [(I) and (J)]. (**K** to **M**) Apex of induced plants showing open structure and straight leaves in confocal images [(L) and (M)]. Scale bars are 5 mm in (A) to (D), 1 mm in (H) and (K), 100 μm in (I) and (J), and 50 μm in (L) and (M).

with the adaxial domain on the inner surface (Fig. 4, K and L, and movie S4). This result suggests that those primordia showing restricted *PHV1* expression at early stages (Fig. 2A) are fated to form traps. The model also predicts that expressing the adaxial domain throughout the primordium eliminates the orthoplanar polarity field, generating a tapering cylinder (Fig. 4B), consistent with the results of ectopic *PHV1* expression (Fig. 3 and fig. S6).

In contrast to *U. gibba*, the adaxial domain in trap primordia of *S. purpurea* extends to the base of the primordium as a narrow ventral strip, which grows out to form a ridge (*5*). Incorporation of such a strip within the above model led to the formation of a cup with a ridge, similar to the form observed in *S. purpurea* (Fig. 4, M to P, and movie S5). In *S. purpurea*, cell divisions were observed to be preferentially periclinal (new cell walls parallel to the outer surface of the primordium) in the ridge-forming region but not in the more distal cup-forming region (*5*). Fukushima *et al.* incorporated these observations into a two-dimensional (2D) model of a transverse section of the primordium, with small outgrowths prespecified at the ad-ab boundary. Assuming divisions are induced near the epidermal ad-ab boundary and that the planes of division determine the orientation of growth, this model could generate the observed 2D shapes (*5*).

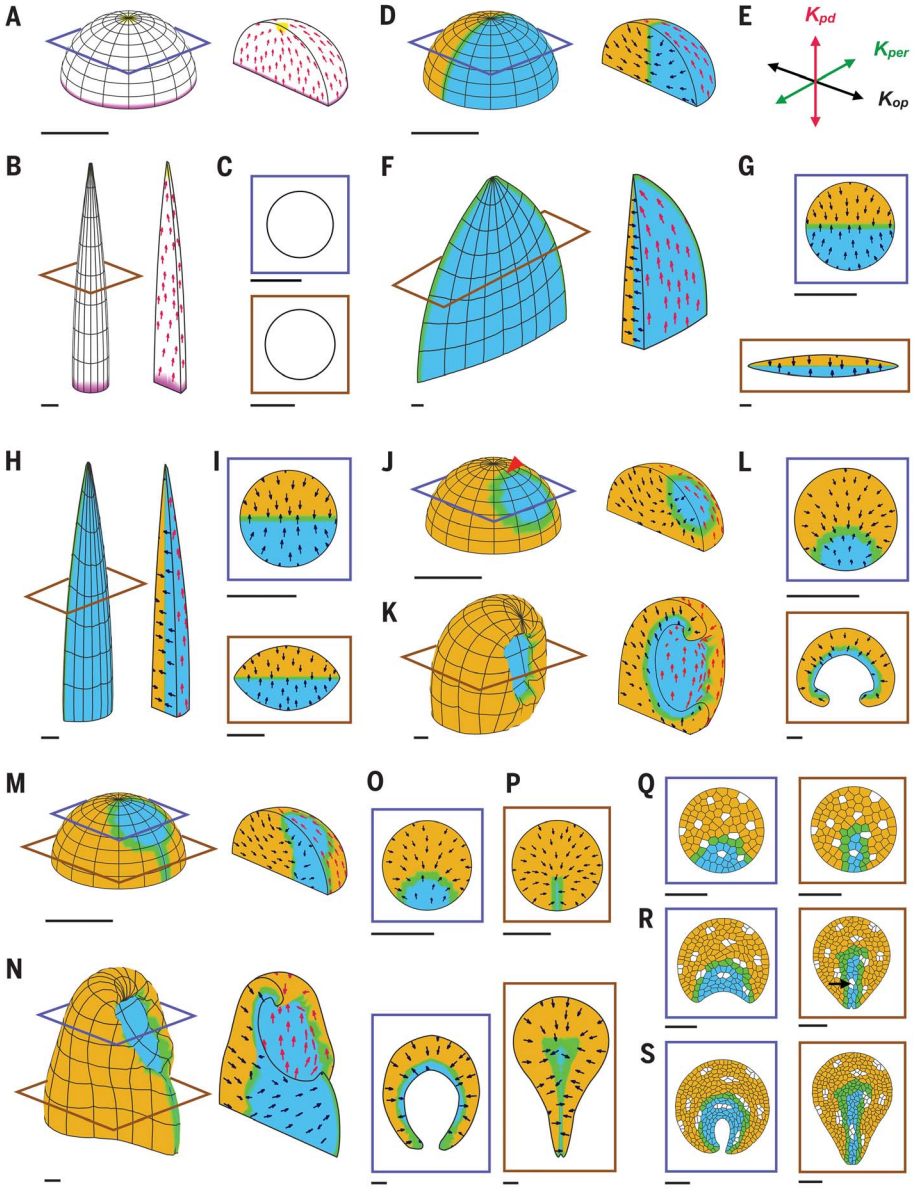


Fig. 4. Modeling of leaf and trap development. Oblique, cutaway, and section views. (**A** to **C**) Generation of tapering cylinder. Initial [(A) and top of (C)] and final [(B) and bottom of (C)] states are shown. Proximodistal polarity (red arrows) runs from organizers at the base (magenta) to the tip (yellow). (**D** to **G**) Generation of flat sheet. Initial [(D) and top of (G)] and final [(F) and bottom of (G)] states are shown. Adaxial (blue), abaxial (brown), and midplane (green) domains are shown. Orthoplanar polarity (black arrows) runs from surface to midplane. Proximodistal and orthoplanar polarity are shown only on outer and cutaway surfaces, respectively. Three growth rates (K_{pd}, K_{op}, and K_{per}) are specified by two polarity fields (E). (**H** and **I**) Generation of an elliptic cylinder. Final state [(H) and bottom of (I)] and initial state [top of (I)]. (**J** to **L**) Generation of a cup. Red arrowhead indicates shifted position of sink for proximodistal polarity. Initial [(J) and top of (L)] and final [(K) and bottom of (L)] states. (**M** to **P**) Generation of cup with ridge. Initial [(M) and top of (O) and (P)] and final [(N) and bottom of (O) and (P)] states. (**Q** to **S**) 2D models showing cell division patterns corresponding to section levels shown in (M) and (N). Initial (Q), intermediate (R), and final (S) states. Periclinal division walls in ridge are indicated by the arrow in (R). Levels of transverse sections are indicated by color-coded rectangles. Scale bars represent the same length in all panels and are in arbitrary units.

To determine whether our model would give similar division patterns, we modeled transverse sections of the primordium containing virtual cells (Fig. 4Q), with specified growth oriented relative to an orthoplanar polarity field. The plane of cell division was set by taking the shortest path through the cell center (*6*, *21*). Some cells were marked in white to allow clones to be visualized. Running this model recapitulated the observed patterns of division (Fig. 4, R and S, and movies S6 and S7), with divisions in the ridge-forming region being preferentially periclinal (black arrow in Fig. 4R). Thus, our model accounts for both trap morphogenesis and observed planes of division in *S. purpurea*.

Our findings provide a simple mechanistic explanation for the generation of planar, filiform, and cup-shaped leaves, through shifts in expression domains or their effects on growth. Unlike previous models, growth is oriented by a polarity field, anchored by ad-ab domains acting throughout the leaf (not just at the epidermal boundary), consistent with observed division patterns. The planes of division are a result (emergent property) of, rather than the cause of, oriented growth. Moreover, our model does not depend on the primordium already having outgrowths in the regions that form the lamina, showing that it can break morphological symmetry rather than simply elaborating it.

Additional structures, such as petioles, can be generated by introducing further domains into the model (figs. S8 to S10 and movies S8 to S10). Diverse shapes and patterns of dissection in the outline of planar leaves may also be generated through modulation of growth oriented by a proximodistal polarity field (*22*–*25*). Thus, a system in which regional identities modify growth rates oriented by two orthogonal polarity fields provides developmental flexibility and can account for how cup-shaped forms evolved multiple times independently from species with planar leaves.

REFERENCES AND NOTES

1. J. R. McConnell, M. K. Barton, *Development* **125**, 2935–2942 (1998).
2. R. Waites, A. Hudson, *Development* **121**, 2143–2154 (1995).
3. A. M. Ellison, N. J. Gotelli, *J. Exp. Bot.* **60**, 19–42 (2009).
4. Y. Hayakawa, M. Tachikawa, A. Mochizuki, *J. Theor. Biol.* **404**, 206–214 (2016).
5. K. Fukushima *et al.*, *Nat. Commun.* **6**, 6450 (2015).

6. S. Fox *et al.*, *PLOS Biol.* **16**, e2005952 (2018).
7. K. J. I. Lee *et al.*, *PLOS Biol.* **17**, e3000427 (2019).
8. R. S. Poethig, I. M. Sussex, *Planta* **165**, 170–184 (1985).
9. E. Ibarra-Laclette *et al.*, *Nature* **498**, 94–98 (2013).
10. T. Lan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E4435–E4441 (2017).
11. J. R. McConnell *et al.*, *Nature* **411**, 709–713 (2001).
12. J. F. Emery *et al.*, *Curr. Biol.* **13**, 1768–1774 (2003).
13. K. R. Siegfried *et al.*, *Development* **126**, 4117–4128 (1999).
14. R. Kennaway, E. Coen, *Open Biol.* **9**, 190057 (2019).
15. K. Abley *et al.*, *Development* **140**, 2061–2074 (2013).
16. C. Mansfield *et al.*, *Curr. Biol.* **28**, 2638–2646.e4 (2018).
17. H. Lin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 366–371 (2013).
18. M. Nakata *et al.*, *Plant Cell* **24**, 519–535 (2012).
19. M. Vandenbussche *et al.*, *Plant Cell* **21**, 2269–2283 (2009).
20. M. J. Scanlon, R. G. Schneeberger, M. Freeling, *Development* **122**, 1683–1691 (1996).
21. L. Errera, *C. R. Acad. Sci.* **103**, 822–824 (1886).
22. T. Blein *et al.*, *Science* **322**, 1835–1839 (2008).
23. A. Hay, M. Tsiantis, *Nat. Genet.* **38**, 942–947 (2006).
24. E. E. Kuchen *et al.*, *Science* **335**, 1092–1096 (2012).
25. A. Richardson, A. B. Rebocho, E. Coen, *Plant Cell* **28**, 2079–2096 (2016).
26. C. D. Whitewoods, E. Coen, *Curr. Biol.* **27**, R910–R918 (2017).

ACKNOWLEDGMENTS

We would like to thank B. and P. Steward at The Fly Trap Plants and T. Bailey from the Carnivorous Plant Society for plants, seeds, and advice. **Funding:** This work was supported by a European Research Council grant (323028-CarnoMorph) and Biotechnology and Biological Sciences Research Council grants (BBS/E/J/000PR9787, BB/M023117/1, and BB/L008920/1) awarded to E.C. C.B. was funded by grant 1243441 from the Biotechnology and Biological Sciences Research Council. B.G. was supported by a Long-Term Fellowship from the Federation of European Biochemical Societies. J.C. was supported by the University of Chinese Academy of Sciences Joint Ph.D. Training Program and a National Natural Science Foundation of China grant (91631308). **Author contributions:** C.D.W. and B.G. performed biological experiments and data analysis and contributed to conceptualization; J.C. performed computational modeling and data analysis and contributed to conceptualization; M.C. and C.P. performed plant transformation; R.K. performed software development and computational modeling; K.L. developed *U. gibba* resources; C.B. performed RNA sequencing and data analysis; M.Y. performed biological experiments and data analysis; and E.C. supervised the research, performed funding acquisition, and contributed to conceptualization. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. Code is available at the following websites: https://github.com/JIC-Enrico-Coen/GFtbox_CarnivorousTraps_2019 and http://cmpdartsvr3.cmp.uea.ac.uk/wiki/BanghamLab/index.php/Software.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6473/91/suppl/DC1
Materials and Methods
Figs. S1 to S10
Tables S1 to S6
References (*27*–*35*)
Movies S1 to S10
Data S1

View/request a protocol for this paper from *Bio-protocol*.

28 June 2019; accepted 12 November 2019
Published online 21 November 2019
10.1126/science.aay5433

EVOLUTIONARY BIOLOGY

Protein-coding changes preceded cis-regulatory gains in a newly evolved transcription circuit

Candace S. Britton[1,2], Trevor R. Sorrells[1,2]*, Alexander D. Johnson[1]†

Changes in both the coding sequence of transcriptional regulators and in the cis-regulatory sequences recognized by these regulators have been implicated in the evolution of transcriptional circuits. However, little is known about how they evolved in concert. We describe an evolutionary pathway in fungi where a new transcriptional circuit (a-specific gene repression by the homeodomain protein Matα2) evolved by coding changes in this ancient regulator, followed millions of years later by cis-regulatory sequence changes in the genes of its future regulon. By analyzing a group of species that has acquired the coding changes but not the cis-regulatory sites, we show that the coding changes became necessary for the regulator's deeply conserved function, thereby poising the regulator to jump-start formation of the new circuit.

Changes in transcriptional circuits over evolutionary time are an important source of organismal novelty. Such circuits are typically composed of one or more transcriptional regulators (sequence-specific DNA binding proteins) and their direct target genes, which contain cis-regulatory sequences recognized by the regulators. Although changes in cis-regulatory sequences are often stressed as sources of novelty that avoid extensive pleiotropy, it is clear that coding changes in the transcriptional regulatory proteins are also of key importance (*1*–*6*). Some well-documented changes in transcriptional circuitry require concerted changes in both elements (*7*, *8*). Although such concerted changes are likely to be widespread, we know little about how they occur.

In this work, we study a case in the fungal lineage where gains in cis-regulatory sequences and coding changes in the transcriptional regulator were both required for a new circuit to have evolved. Specifically, we addressed which came first: the changes in the regulatory protein or the changes in the cis-regulatory sequences of its 5 to 10 target genes. The system we analyzed consists of an ancient regulator, the homeodomain protein Matα2, and the changes—both in the protein itself and in the regulatory regions of the genes it controls—that occurred across the Saccharomycotina clade of fungi, which spans roughly 300 million years. [In terms of protein diversity, this represents roughly the range between humans and sea sponges (*9*)]. Throughout this time, Matα2 has maintained its ancient function: It binds cooperatively to DNA with a second homeodomain protein, Mata1, to repress a group of genes called the haploid-specific genes (Fig. 1). More recently, Matα2 formed an additional circuit, which is present in only a subset of the Saccharomycotina: It binds DNA cooperatively with the MADS box protein Mcm1 to repress the a-specific genes (Fig. 1). Before this time, the a-specific genes were

[1]Department of Microbiology and Immunology, University of California, San Francisco, CA 94158, USA. [2]Tetrad Graduate Program, University of California, San Francisco, CA 94158, USA.
*Present address: Laboratory of Neurogenetics and Behavior, The Rockefeller University, New York, NY 10065, USA.
†Corresponding author. Email: ajohnson@cgl.ucsf.edu

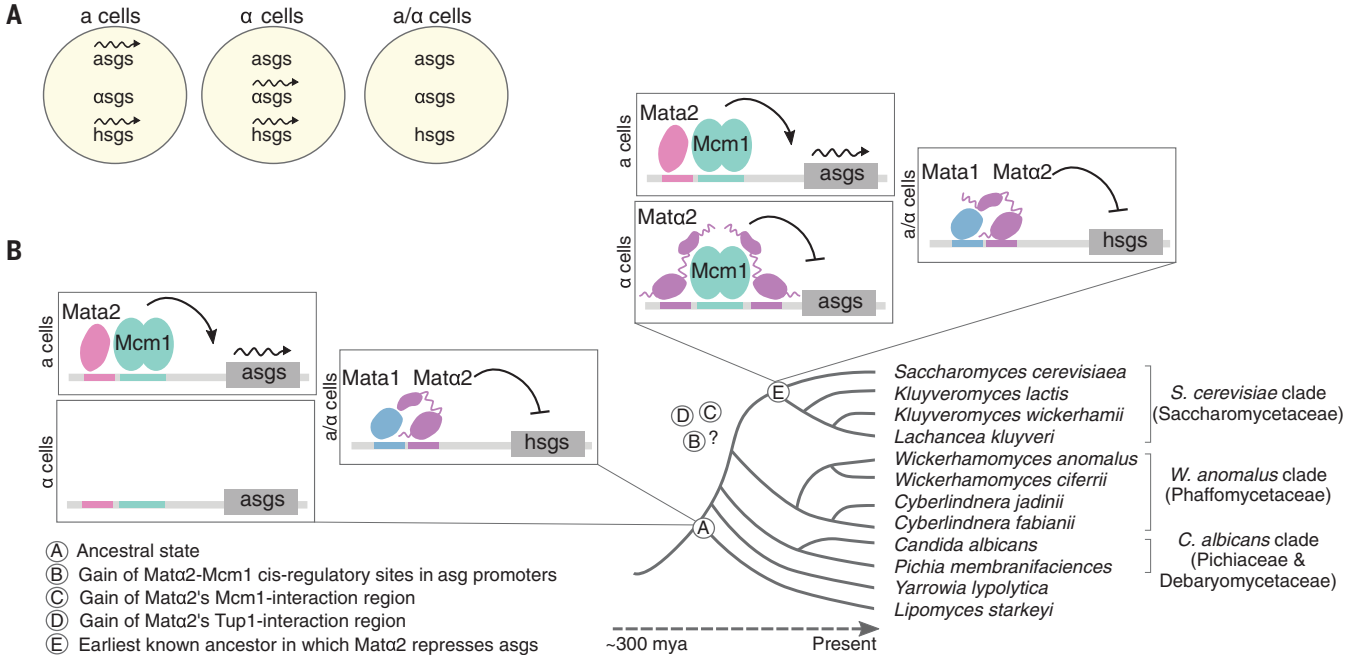
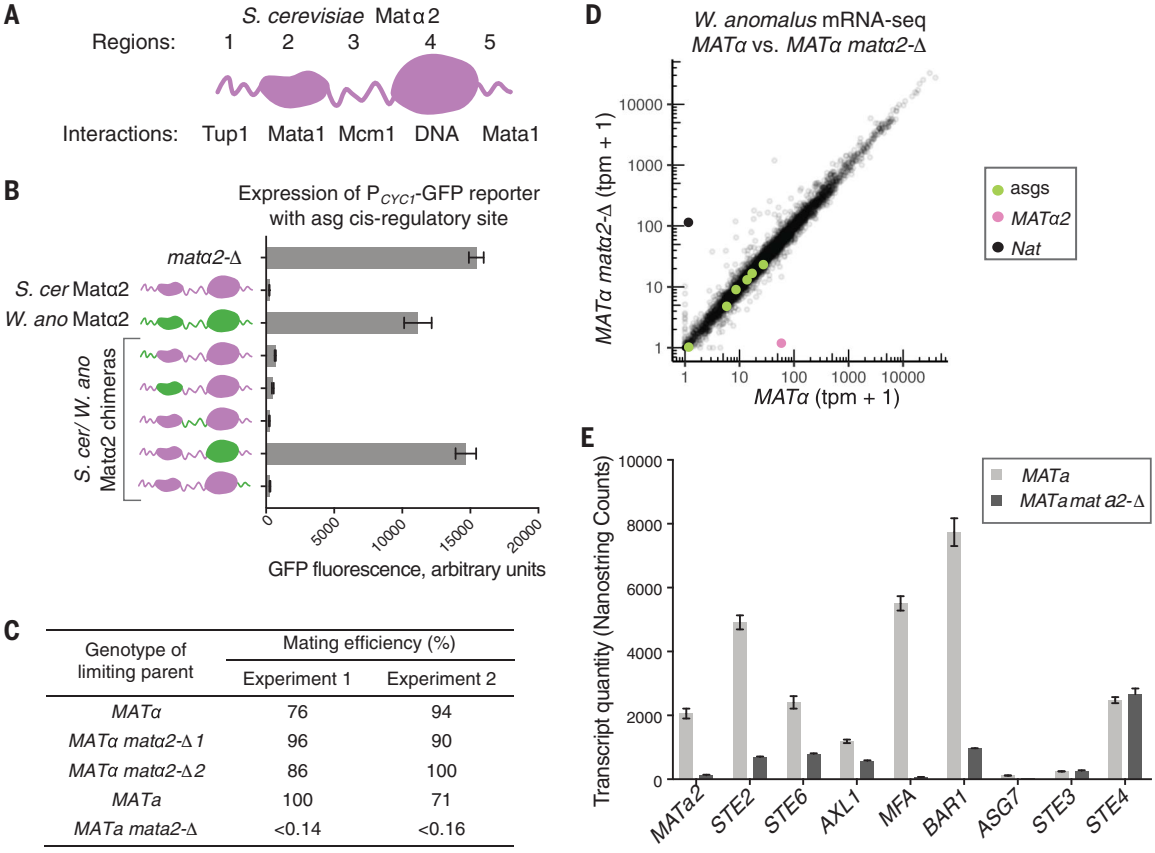


Fig. 1. Cell type–specific gene expression in the Saccharomycotina yeast. (**A**) Across the Saccharomycotina clade, a and α cells each express a set of genes specific to that cell type (a- and α-specific genes, or asgs and αsgs, respectively), as well as a shared set of haploid-specific genes (hsgs). a and α cells can mate to form a/α cells, which do not express the a-, α-, or haploid-specific genes (*22*). Wavy arrows represent active transcription. (**B**) The mechanism underlying the expression of a-specific genes is different among species. In the last common ancestor of the Saccharomycotina yeast (see circled A in the figure), transcription of the a-specific genes was activated by Mata2, a protein produced only in a cells, which binds directly to the regulatory region of each a-specific gene (*10*, *23*). Much later in evolutionary time (see circled E in the figure), repression of the a-specific genes by direct binding by Matα2 evolved. Still later, the Mata2-positive form of control was lost in some species (including *S. cerevisiae*), leaving only the Matα2-negative form. mya, million years ago.



C

Genotype of limiting parent	Mating efficiency (%)	
	Experiment 1	Experiment 2
MATα	76	94
MATα matα2-Δ1	96	90
MATα matα2-Δ2	86	100
MATa	100	71
MATa mata2-Δ	<0.14	<0.16

Fig. 2. *W. anomalus* Matα2 has functional Tup1- and Mcm1-interacting regions but does not repress the a-specific genes. (**A**) The five modules of the *S. cerevisiae* Matα2 protein. Structural domains are shown as globular, and unstructured regions are shown as wavy lines. (**B**) Expression of an a-specific gene reporter in the presence of *S. cerevisiae* (*S. cer*) Matα2 (purple), *W. anomalus* (*W. ano*) Matα2 (green), and hybrid proteins (purple and green). Means and SDs of three independent genetic isolates, grown and tested in parallel, are shown. GFP, green fluorescent protein. (**C**) In *W. anomalus*, Mata2, but not Matα2, is required for a cells to mate (see supplementary text for details). (**D**) mRNA sequencing (mRNA-seq) (tpm, transcripts per million) of wild-type *W. anomalus* α cells (*MAT*α) compared with α cells with *MAT*α*2* deleted (*MAT*α*2 matα2-*Δ). a-specific genes *STE2*, *AXL1*, *ASG7*, *BAR1*, *STE6*, and *MATa2* are shown in green. Expression of *MAT*α*2* and the marker used to delete it (*Nat*) are shown in pink and opaque black, respectively. Data from independent replicates are given in fig. S3. (**E**) a-specific gene expression levels in a wild-type *W. anomalus* a cells (*MATa*) compared with a cells with *MATa2* deleted (*MATa2 mata2-*Δ), measured by the NanoString nCounter system (*24*). For comparison, expression levels of the α-specific gene *STE3* and the haploid-specific gene *STE4* are also given. Means and SDs of two cultures per genotype, grown and tested in parallel, are shown.

regulated by a different mechanism—positive control by the HMG-domain protein Mata2 (*10*, *11*).

The switch between the two mechanisms of controlling the a-specific genes occurred sometime before the divergence of *Saccharomyces cerevisiae* and *Kluyveromyces lactis* (formally known as the Saccharomycetaceae, here called the *S. cerevisiae* clade) but after the divergence of this clade and that containing *Candida albicans* and *Pichia membrifaciens* (formally known as the Pichiaceae and Debaryomecetaceae, here called the *C. albicans* clade) (Fig. 1B). Three events must have occurred for the newer (repression) scheme to have evolved: (i) Matα2 acquired the ability to contact the Tup1-Ssn6 co-repressor, bringing it to DNA to carry out the repression function; (ii) Matα2 acquired the ability to bind to DNA cooperatively (through a direct protein-protein contact) with Mcm1; and (iii) the a-specific genes (numbering between 5 and 10, depending on the species) each acquired a new cis-regulatory site for the Matα2-Mcm1 combination (Fig. 1B).

To determine the order of these events, we studied Matα2 and the regulation of the a-specific genes in a clade that branched from the ancestor before the occurrences of all three of these events. We reasoned that this group of species might have acquired some, but not all, of the changes needed to form the new circuit, and it therefore might provide clues to the evolutionary history. This approach was made possible by the genome sequencing of a monophyletic group of species that branches before the last common ancestor of the *S. cerevisiae* clade (formally known as the Phaffomycetaceae) (Fig. 1B) (*12*, *13*). We chose the species *Wickerhamomyces anomalus*, and we were able to optimize relatively simple procedures to alter it genetically (*14*).

We examined the *W. anomalus* Matα2 protein sequence to determine whether it is more similar to the ancestral (represented by *C. albicans*) or the derived (represented by *S. cerevisiae*) form of Matα2. Alignment of the Matα2 coding sequences across many species indicated that, of the five functional regions described for the *S. cerevisiae* protein (Fig. 2A and fig. S1), the *W. anomalus* protein shares all of them. In particular, it has a similar Tup1-interacting region (region 1, Fig. 2A) and Mcm1-interacting region (region 3, Fig. 2A); these regions are missing in outgroup proteins and are needed to repress the a-specific genes in *S. cerevisiae* (*11*, *15*). By swapping these *W. anomalus* regions into the *S. cerevisiae* protein, we confirmed that they are functional in repressing the a-specific genes (Fig. 2B). In the course of these experiments, we found that the homeodomain of the *W. anomalus* protein contained mutations that prevented its binding to the a-specific gene cis-regulatory sequence in *S. cerevisiae*, a derived change within this clade alone (Fig. 2B and fig. S1). Similar results were obtained with the Matα2 protein from two additional species that branch with *W. anomalus*, indicating that these two conclusions—that *W. anomalus* clade Matα2 bears functional protein-protein interactions but cannot bind

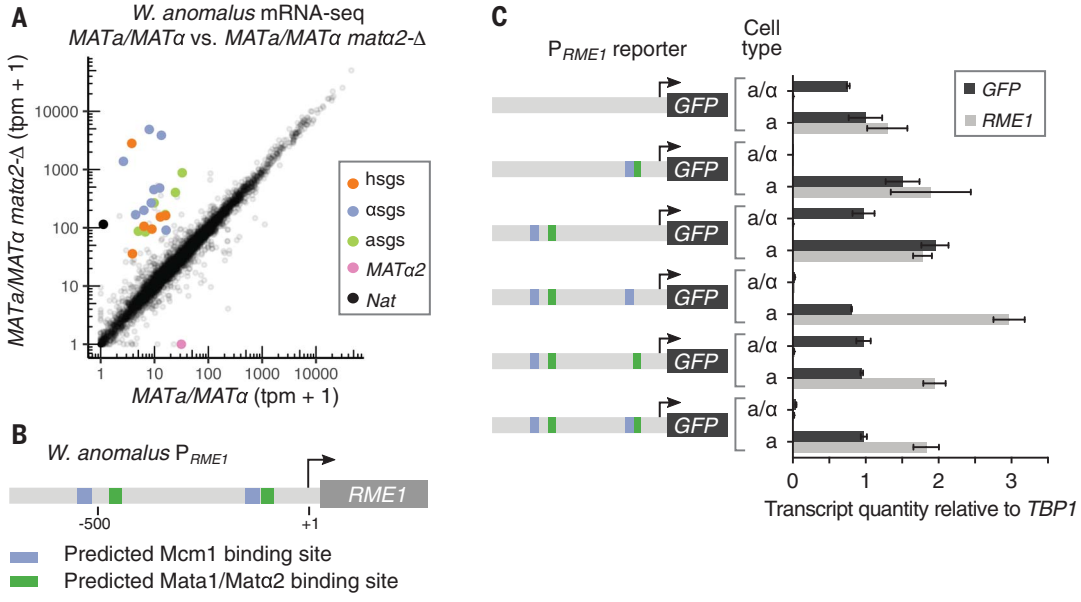


Fig. 3. Mata1, Matα2, and an Mcm1 cis-regulatory sequence are all required for haploid-specific gene repression in *W. anomalus*. (**A**) mRNA-seq of a wild-type *W. anomalus* a/α cell (*MATa/MATα*) compared with an a/α cell with *MATα2* deleted (*MATa/MATα matα2*-Δ). The a-specific genes are shown in green, the haploid-specific genes in orange, and the α-specific genes in blue. Data from one culture of each genotype are plotted here, and data from replicates, grown and prepared in parallel, and similar results obtained by deleting Mata1 are shown in fig. S5. (**B**) Diagram of the sequence upstream of the *RME1* coding sequence indicating presumptive Mata1-Matα2 (green) and Mcm1 (blue) binding sites. Arrow indicates the transcription start site. (**C**) Expression levels of endogenous *RME1* transcript (which serves as a control) and various P_{RME1}-GFP reporter constructs in *W. anomalus* a and a/α cells measured by reverse transcription quantitative polymerase chain reaction. Quantities are means and SDs of two cultures grown and measured in parallel, normalized to expression of the housekeeping gene *TBP1*. Independent replicates are given in fig. S6.

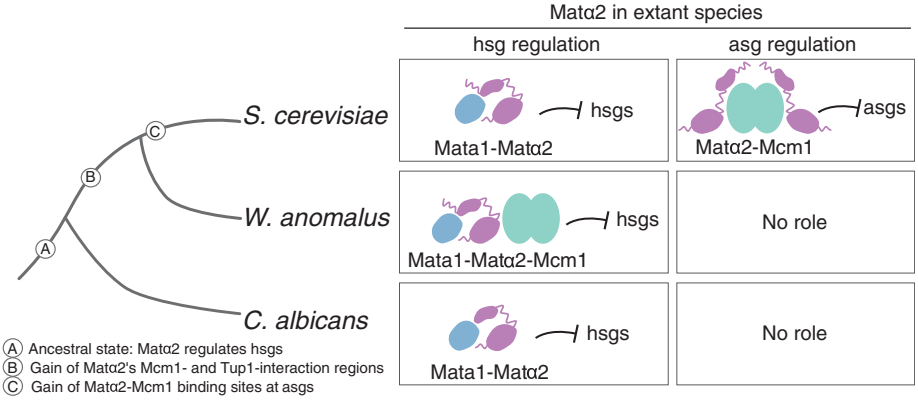


Fig. 4. Order of evolutionary events leading to repression of the a-specific genes by Matα2. The three-protein solution for repressing the haploid-specific genes remains in the *W. anomalus* clade, but in the *S. cerevisiae* lineage it was partitioned into a-specific gene regulation (which uses only two proteins, Mcm1 and Matα2) and repression of the haploid-specific genes (which requires Matα2 and Mata1). The three-protein intermediate explains how the necessary changes in the regulatory protein Matα2 could have been maintained for millions of years before being co-opted for the new circuit.

the *S. cerevisiae* a-specific genes—are characteristic of the *W. anomalus* clade rather than of a single species (fig. S1D).

The observation that the *W. anomalus* Matα2 protein acquired the necessary coding changes to interact with Tup1 and Mcm1 but could not bind to the *S. cerevisiae* a-specific gene control region raised the question of whether it has any role in regulating the a-specific genes in *W. anomalus*. A series of otherwise-isogenic strains was constructed with Matα2 (and Mata2) deleted, and the results show that, in this species, Matα2 does not regulate the a-specific genes; they are instead regulated by Mata2 (Fig. 2, C to E, and fig. S3). Thus, despite the changes in Matα2, *W. anomalus* retains the ancestral form of a-specific gene regulation and activation by Mata2. This conclusion is supported by a bioinformatic analysis showing that the a-specific genes possess Mata2-Mcm1, but not Matα2-Mcm1 cis-regulatory sequences (fig. S4B). These results argue against the possibility that direct, a-specific gene repression by Matα2 existed in an ancestor of *W. anomalus* but was subsequently lost, as this would have required the independent loss of Matα2 binding sites from all of the a-specific genes across numerous species.

Our experiments up to this point demonstrate that Matα2 had acquired the coding changes needed to repress the a-specific genes millions of years before its cis-regulatory sequences appeared in the a-specific genes. We next addressed how these changes in the Matα2 protein could have been maintained in the absence of their usefulness in repressing the a-specific genes. One hypothesis focuses on Matα2's ancient function—repressing the haploid-specific genes with Mata1—and holds that the Matα2 coding changes became required for this function only in the *W. anomalus* clade. To test this idea, we analyzed the requirements for haploid-specific gene repression in *W. anomalus*. We deleted *MATα2* and *MATa1* in a/α cells and found that they are both necessary for haploid-specific gene repression, a conclusion confirmed by chromatin immunoprecipitation (Fig. 3A and figs. S5 and S6C). However, unlike in species outside the *W. anomalus* clade, the Tup1-interaction region and the Mcm1-interaction region of Matα2 are necessary for repression of the haploid-specific genes within the clade (Fig. 2A and fig. S6B). Finally, an Mcm1 cis-regulatory site is also required for the repression of the *W. anomalus* haploid-specific gene *RME1* (Fig. 3C and fig. S6). Taken together, these experiments show that Matα2, Mata1, and Mcm1 are all required for haploid-specific gene repression in *W. anomalus*, and that the portions of Matα2 that interact with Mcm1 and Tup1 are

also required. This three-part recognition of the haploid-specific genes in the *W. anomalus* clade was not anticipated from studies of other species. Even in the *S. cerevisiae* clade, where Mcm1 and Matα2 are known to interact, this interaction is not required for haploid-specific gene repression (*11*). These results explain the observation that the key changes in Matα2 needed for the new a-specific gene circuit were already in place in the last common ancestor of *S. cerevisiae* and *W. anomalus*, long before the circuit came into play (Fig. 4). An alternative scenario—in which the Matα2 protein gained the Mcm1-interaction region twice, once in the *S. cerevisiae* clade and once in the *W. anomalus* clade—is unlikely because the same seven amino acids would have had to be gained in exactly the same position in the protein (fig. S1).

This study helps to illuminate several long-standing issues. First, how is pleiotropy avoided when transcriptional regulators acquire new functions? The modular structure of Matα2 is evident from the protein domain swap experiments (Fig. 2B and fig. S6B), showing that the derived regions of the protein (Tup1- and Mcm1-interaction regions) can be transplanted to a variety of outgroup Matα2 proteins and that they endow the ancestral proteins with the new functions without compromising the existing functions (*11*). However, there is a second, more subtle way that extensive pleiotropy was avoided in the case studied in this work. In the shift between the different ways of controlling the haploid-specific genes, pleiotropy was avoided automatically; even before the new a-specific gene circuit was formed, the Matα2-Mcm1 combination (which forms the basis of the new circuit) had been "vetted" for millions of years as being compatible with the ancestral function of Matα2.

Second, is the evolutionary pathway we describe in this paper compatible with the concept of constructive neutral evolution, or the idea that new functions can evolve through evolutionary transitions of approximately equal fitness (*16*–*18*)? Before the results presented here were obtained, it was difficult to understand how the derived circuit represented by *S. cerevisiae* (repression of the a-specific genes by Matα2 in α cells) could have evolved because it required changes in both the Matα2 coding region and in the cis-regulatory sequences controlling the 5 to 10 a-specific genes. We propose that the prior changes to Matα2 represent an example of constructive neutral evolution, in the sense that the neutral sampling of different ways to repress the haploid-specific genes over evolutionary time led to changes in Matα2 that, millions of years later through exaptation, formed the basis of the new circuit. Although we cannot rule out the possibility that the differences in the way that the haploid-specific genes were repressed were somehow adaptive, it seems more likely that they occurred neutrally—an explanation consistent with a wide variety of theoretical work (*16*–*19*). In any case, there is no obvious adaptive explanation, and neutral evolution is an appropriate default hypothesis.

Third, is there an inherent logic to the mechanisms underlying a given transcription circuit? In this paper, we show that some clades regulate the haploid-specific genes with a combination of three proteins, whereas others use only two of the proteins, even though the third is present. Nonetheless, the overall pattern of haploid-specific gene expression is the same. If there is any overriding design logic to the different mechanisms of regulating these genes, it is difficult to discern (*20*). More broadly, the work presented here illustrates that a given transcription circuit is best understood as one of several possible interchangeable, mechanistic solutions rather than as a finished, optimized design (*21*).

REFERENCES AND NOTES

1. V. J. Lynch, G. P. Wagner, *Evolution* **62**, 2131–2154 (2008).
2. A. M. Cheatle Jarvela, V. F. Hinman, *EvoDevo* **6**, 3 (2015).
3. D. L. Stern, V. Orgogozo, *Evolution* **62**, 2155–2177 (2008).
4. G. A. Wray, *Nat. Rev. Genet.* **8**, 206–216 (2007).
5. P. J. Wittkopp, G. Kalay, *Nat. Rev. Genet.* **13**, 59–69 (2011).
6. H. Li, A. D. Johnson, *Curr. Biol.* **20**, R746–R753 (2010).
7. C. Sayou *et al*., *Science* **343**, 645–648 (2014).
8. C. R. Baker, B. B. Tuch, A. D. Johnson, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7493–7498 (2011).
9. X.-X. Shen *et al*., *Cell* **175**, 1533–1545.e20 (2018).
10. A. E. Tsong, M. G. Miller, R. M. Raisner, A. D. Johnson, *Cell* **115**, 389–399 (2003).
11. C. R. Baker, L. N. Booth, T. R. Sorrells, A. D. Johnson, *Cell* **151**, 80–95 (2012).
12. X.-X. Shen *et al*., *G3* **6**, 3927–3939 (2016).
13. R. Riley *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **113**, 9882–9887 (2016).
14. C. Kurtzman, J. W. Fell, Eds., *The Yeasts - A Taxonomic Study* (Elsevier, 1998).
15. K. Komachi, A. D. Johnson, *Mol. Cell. Biol.* **17**, 6023–6028 (1997).
16. A. Stoltzfus, *J. Mol. Evol.* **49**, 169–181 (1999).
17. J. Lukeš, J. M. Archibald, P. J. Keeling, W. F. Doolittle, M. W. Gray, *IUBMB Life* **63**, 528–537 (2011).
18. M. W. Gray, J. Lukes, J. M. Archibald, P. J. Keeling, W. F. Doolittle, *Science* **330**, 920–921 (2010).
19. A. Wagner, *FEBS Lett.* **579**, 1772–1778 (2005).
20. C. K. Dalal, A. D. Johnson, *Genes Dev.* **31**, 1397–1405 (2017).
21. T. R. Sorrells, A. D. Johnson, *Cell* **161**, 714–723 (2015).
22. I. Herskowitz, *Nature* **342**, 749–757 (1989).
23. A. E. Tsong, B. B. Tuch, H. Li, A. D. Johnson, *Nature* **443**, 415–420 (2006).
24. G. K. Geiss *et al*., *Nat. Biotechnol.* **26**, 317–325 (2008).

ACKNOWLEDGMENTS

We thank L. Noiman, M. Lohse, C. Dalal, K. Fowler, and L. Booth for comments on the manuscript and C. Baker, I. Nocedal, N. Ziv, and B. Heineke for advice. We thank C. Schorsch of Evonik Industries for providing us with the plasmid used to genetically modify *W. anomalus*. **Funding:** The work was supported by NIH grant R01 GM037049 (to A.D.J.), an ARCS Scholarship (to C.S.B.), and an NSF Graduate Fellowship (to T.R.S.). **Author contributions:** C.S.B., T.R.S., and A.D.J. designed and interpreted experiments and wrote and edited the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Plasmid pCS.ΔLig4 can be obtained from Evonik Industries under a material transfer agreement. mRNA-seq data have been deposited at the National Center for Biotechnology Information Gene Expression Omnibus under accession number GSE133191.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6473/96/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S6
References (*25*–*45*)
Data S1

View/request a protocol for this paper from *Bio-protocol*.

1 April 2019; accepted 5 November 2019
10.1126/science.aax5217

TUBULIN

TTC5 mediates autoregulation of tubulin via mRNA degradation

Zhewang Lin[1], Ivana Gasic[2]*, Viswanathan Chandrasekaran[1]*, Niklas Peters[1]†, Sichen Shao[3], Timothy J. Mitchison[2], Ramanujan S. Hegde[1]‡

Tubulins play crucial roles in cell division, intracellular traffic, and cell shape. Tubulin concentration is autoregulated by feedback control of messenger RNA (mRNA) degradation via an unknown mechanism. We identified tetratricopeptide protein 5 (TTC5) as a tubulin-specific ribosome-associating factor that triggers cotranslational degradation of tubulin mRNAs in response to excess soluble tubulin. Structural analysis revealed that TTC5 binds near the ribosome exit tunnel and engages the amino terminus of nascent tubulins. TTC5 mutants incapable of ribosome or nascent tubulin interaction abolished tubulin autoregulation and showed chromosome segregation defects during mitosis. Our findings show how a subset of mRNAs can be targeted for coordinated degradation by a specificity factor that recognizes the nascent polypeptides they encode.

Alpha and beta tubulins form obligate heterodimers (hereafter αβ-tubulin) that reversibly and dynamically polymerize into microtubules—cytoskeletal elements that regulate cell shape, drive mitosis, provide platforms for intracellular transport, and mediate cell movement (*1*). Microtubule dynamics, and the various processes that depend on it (*2*, *3*), is strongly influenced by the concentration of soluble (i.e., nonpolymerized) αβ-tubulin (*4*). When cells detect an increase in soluble αβ-tubulin concentration, they trigger degradation of tubulin mRNAs via a process termed tubulin autoregulation (*5*–*7*).

Autoregulation requires translation, indicating that ribosome-engaged tubulin mRNAs are selectively targeted for degradation (*8*, *9*). Analysis of β-tubulin autoregulation in mammalian cells indicates a critical role for the first four residues (Met-Arg-Glu-Ile, or MREI) common to all β-tubulin isoforms (*10*, *11*). Because autoregulation is prevented by physical occlusion of the MREI motif (*12*), a factor is thought to engage this sequence on nascent tubulin to initiate degradation of the mRNA being translated.

We used a site-specific photo–cross-linking strategy (Fig. 1A and fig. S1) to detect cytosolic factors that specifically recognize the N-terminal autoregulatory motif (MREI) of nascent β-tubulin early during its translation. A ribosome–nascent chain complex (RNC) displaying the first 94 amino acids of [^{35}S]methionine-labeled human β-tubulin containing the ultraviolet light (UV)–activated cross-linking amino acid *p*-benzoyl-L-phenylalanine (Bpa) was produced by in vitro translation in rabbit reticulocyte lysate. Irradiation of these RNCs with UV light generated nascent chain cross-links to various proteins, only one of which was sensitive to mutation of residues 2, 3, and 4 of the MREI motif (Fig. 1B). This MREI-specific interaction partner was identified by quantitative mass spectrometry to be TTC5 (Fig. 1C), a highly conserved protein found widely across eukaryotes (fig. S2). TTC5 engaged the MREC motif at the N terminus of nascent α-tubulin comparably to the MREI motif on β-tubulin (Fig. 1D), consistent with position 4 being less critical than positions 2 or 3 (Fig. 1B) (*11*). Thus, TTC5 is a nascent polypeptide binding protein specific for the N termini of α- and β-tubulins.

To understand how TTC5 engages its substrates on the ribosome, we purified nascent tubulin RNCs in complex with TTC5 (fig. S3) and determined the structure of this complex by single-particle cryo–electron microscopy (cryo-EM). The TTC5-RNC reconstruction (figs. S4 and S5) showed the ribosome with a peptidyl-tRNA, a nascent β-tubulin polypeptide within the ribosome exit tunnel, and TTC5 bound at the mouth of the tunnel (Fig. 2A). The heterodimeric nascent polypeptide–associated complex (NAC) was observed at its previously established binding site (*13*, *14*) opposite the exit tunnel from TTC5 (see fig. S4). NAC is not specific to tubulin RNCs (*15*, *16*), does not contact TTC5 in the structure, and is not discussed further.

TTC5 was seen to make two contacts with the ribosome. The first contact involves three highly conserved lysine side chains in the oligonucleotide-binding domain of TTC5 making electrostatic interactions with phosphates of the 28*S* rRNA backbone (Fig. 2B). The second contact involves ribosomal protein uL24 and buries ~500 Å^2 of TTC5 adjacent to a deep groove formed by the tetratricopeptide repeat domain of TTC5 (Fig. 2C). The groove faces the mouth of the exit tunnel and contains cryo-EM density that we assigned to the first eight amino acids of β-tubulin (fig. S5), consistent with photo–cross-linking results (fig. S6).

The structural model allowed us to deduce likely interactions between the MREI motif and conserved side chains lining the TTC5 groove (Fig. 2D). Depending on its orientation, Arg^2 of nascent tubulin is within salt-bridge distance of Glu^{259} and Asp^{225} of TTC5. Glu^3 in nascent tubulin would likely interact with Arg^{147} in TTC5. Ile^4 faces a moderately hydrophobic surface that could accommodate a cysteine (as in α-tubulins) or possibly other amino acids, consistent with earlier mutagenesis (*11*). Collectively, the structure shows how TTC5 binds near the ribosome exit tunnel with its peptide-binding groove positioned to engage nascent tubulins shortly after they emerge from the ribosome.

Recombinant TTC5 containing Bpa at position 194 in the "floor" of the peptide binding groove (Fig. 2D) efficiently cross-linked with MREI-containing nascent chains, weakly cross-linked with MREV-containing nascent chains, and did not form cross-links with any other mutants (Fig. 3A). Analysis of RNC cross-linking with various TTC5 mutants (Fig. 3B) validated Arg^{147}, Asp^{225}, and Glu^{259} as key residues within the groove that likely interact with Arg^2 and Glu^3 of nascent tubulin (see Fig. 2D). Binding assays with purified TTC5 and synthetic peptides (Fig. 3C and fig. S7) verified these findings and additionally showed that Met^1 of nascent tubulin is critical for TTC5 binding and must strictly be at the N terminus. Thus, the structural analysis rationalizes all earlier β-tubulin mutagenesis studies on autoregulation requirement (*11*) and reveals the mechanistic basis of the exquisite specificity of autoregulation for α- and β-tubulins (*5*) that uniquely contain an MREI or MREC motif at the N terminus (*17*).

Mutating the ribosome-interacting residues Lys^{285} and Lys^{287} of TTC5 to glutamic acid (KK-EE) completely abolished β-tubulin RNC binding in the cross-linking assay (Fig. 3B), despite unperturbed binding of TTC5 to synthetic tubulin autoregulatory peptide in a thermal shift assay (Fig. 3C and fig. S7). Affinity purification of recombinant TTC5 from in vitro translation reactions of nascent β-tubulin RNCs showed that no ribosomes were recovered with either TTC5(KK-EE) or the peptide-binding mutant TTC5(R147A), in contrast to wild-type TTC5 (Fig. 3D). Thus, the avidity of bipartite binding to the ribosome and nascent tubulin imparts high affinity and specificity to the TTC5-RNC interaction.

CRISPR-mediated disruption of TTC5 expression in multiple cell lines completely abolished the decay of α- and β-tubulin mRNAs in

[1]MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK. [2]Department of Systems Biology, Blavatnik Institute, Harvard Medical School, Boston, MA 02115, USA. [3]Department of Cell Biology, Blavatnik Institute, Harvard Medical School, Boston, MA 02115, USA.
*These authors contributed equally to this work. †Present address: Center for Molecular Biology of Heidelberg University (ZMBH), 69120 Heidelberg, Germany.
‡Corresponding author. Email: rhegde@mrc-lmb.cam.ac.uk

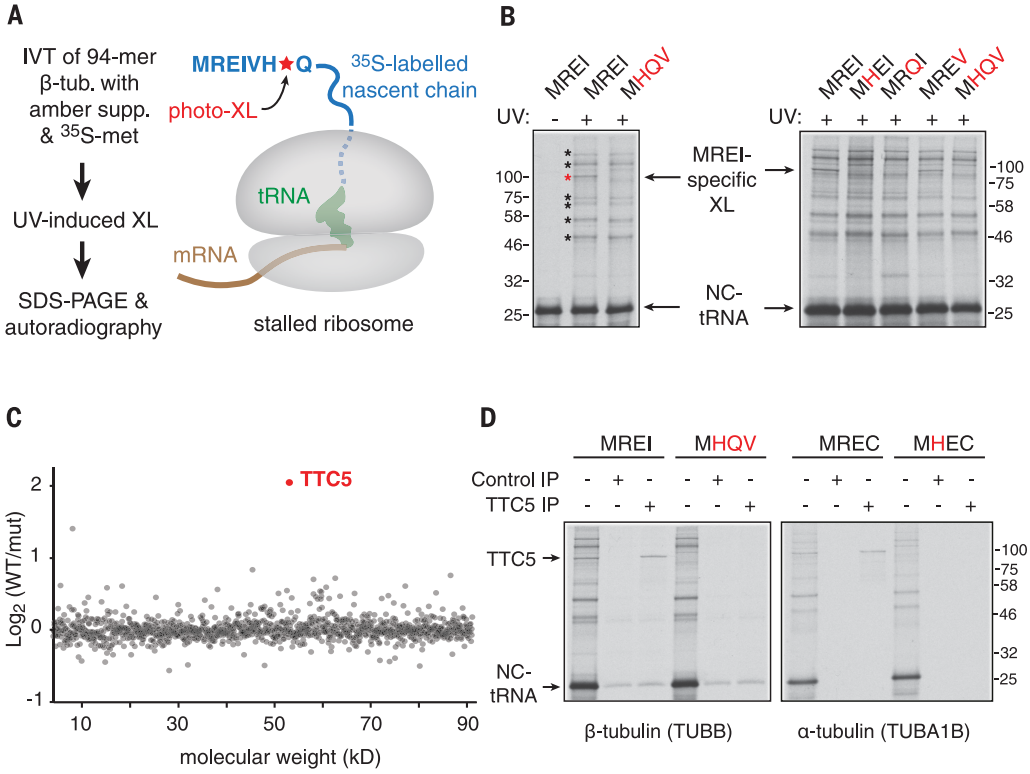


Fig. 1. TTC5 interacts with the N termini of nascent tubulins. (**A**) Experimental strategy to detect interaction partners close to the N terminus of nascent tubulin. The UV-activated cross-linking amino acid *p*-benzoyl-L-phenylalanine (Bpa) is introduced site-specifically at position 7 using amber suppression (see fig. S1). IVT, in vitro translation; XL, cross-linking; PAGE, polyacrylamide gel electrophoresis. (**B**) Photo–cross-linking analysis of ^{35}S-labeled 94–amino acid ribosome–nascent chain complexes (RNCs) of human β-tubulin and mutants (indicated in red) in the N-terminal MREI motif. The positions of non–cross-linked tRNA-associated nascent chain (NC-tRNA) and a cross-linking partner specific to wild-type tubulin (red asterisk) are indicated. Other nascent chain cross-links agnostic to the MREI motif are indicated by black asterisks. (**C**) Quantitative mass spectrometry of proteins copurified with wild-type (WT) versus MHQV mutant (mut) β-tubulin RNCs plotted by molecular weight. (**D**) Photo–cross-linking and immunoprecipitation (IP) analysis of ^{35}S-labeled 94–amino acid RNCs of α- or β-tubulin compared to the indicated mutants. Amino acid abbreviations in this or later figures: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; M, Met; Q, Gln; R, Arg; V, Val.

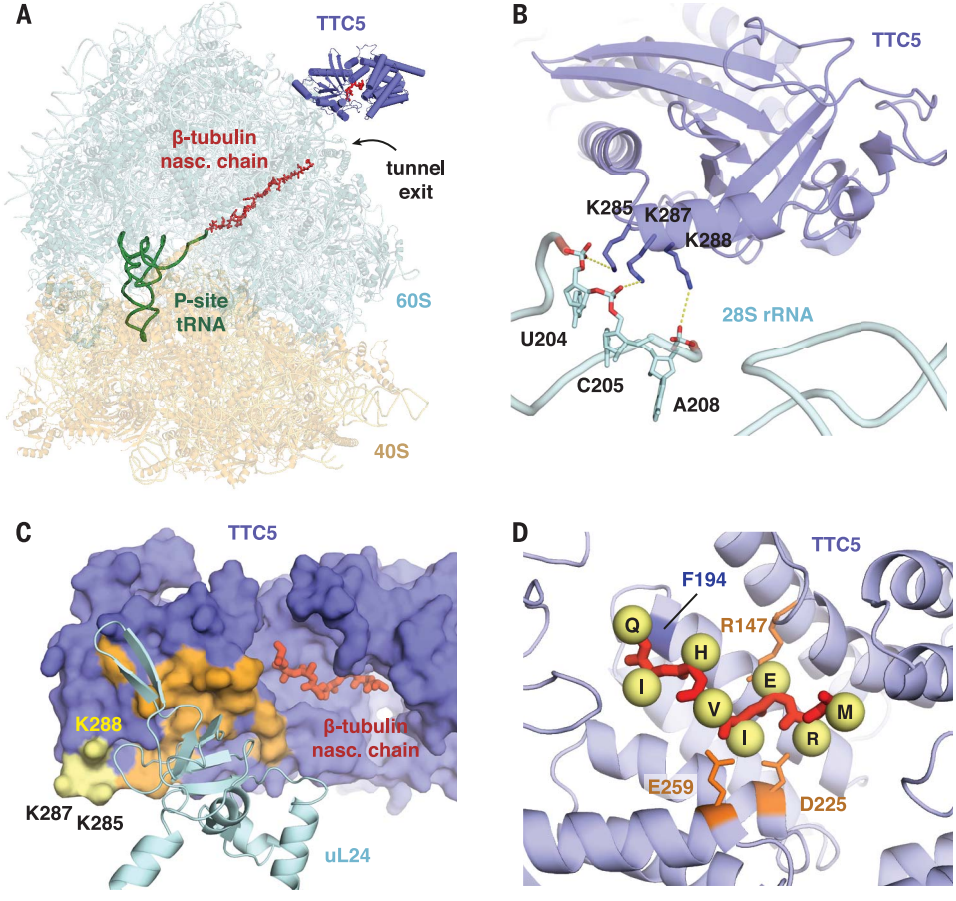


Fig. 2. Mechanism of ribosome–nascent chain engagement by TTC5. (**A**) Overview of the cryo-EM–derived structure of the complex between TTC5 and a ribosome containing the first 64 amino acids of β-tubulin. (**B**) Close-up view of the TTC5 interaction with 28*S* rRNA. Three conserved Lys residues in TTC5 within electrostatic interaction distance of the rRNA backbone are indicated. (**C**) The surface of TTC5 that interacts with ribosomal protein uL24 is indicated in orange. The 28*S*-interacting residues from (B) are shown in yellow. The N-terminal eight amino acids of the β-tubulin nascent chain are shown in red within the chain's binding groove of TTC5. (**D**) Close-up view of the N-terminal eight amino acids of nascent β-tubulin (MREIVHIQ) within TTC5. Yellow spheres denote C-β atoms for the indicated side chains (not modeled). Arg^{147} is within salt-bridge distance of Glu^{3}, and Asp^{225} and Glu^{259} are within salt-bridge distance of Arg^{2}. Phe^{194} on the "floor" of the binding groove, shown in Fig. 3A to cross-link with nascent β-tubulin, is indicated.

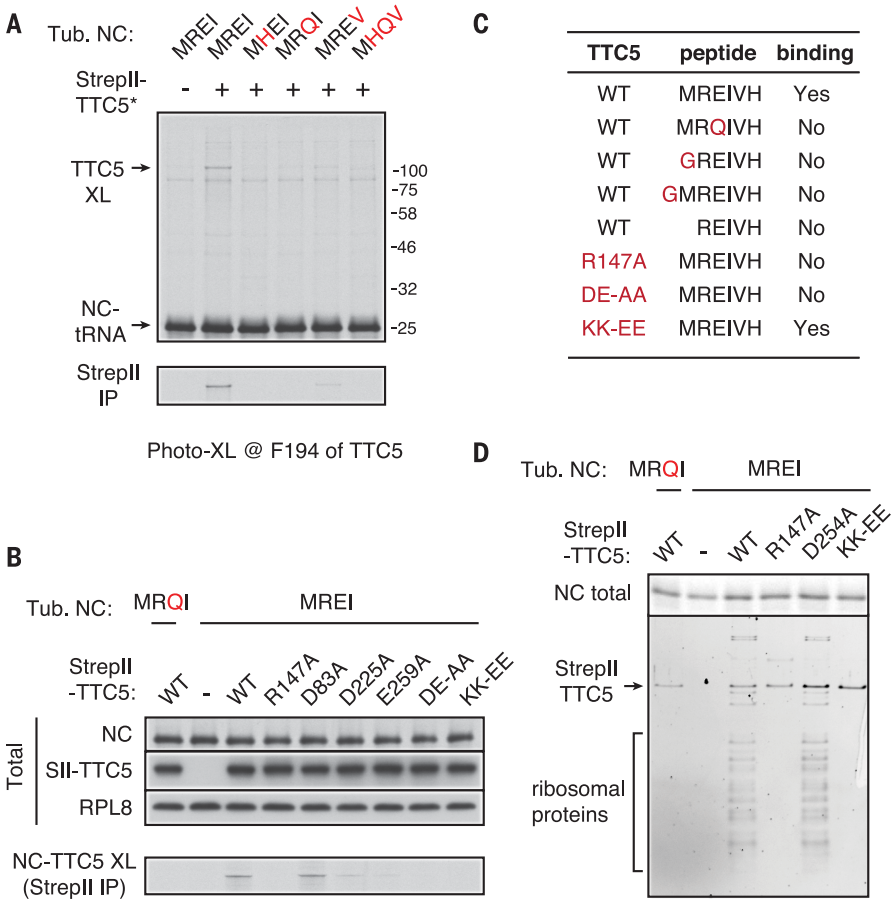


Fig. 3. Avidity-based RNC binding imparts specificity to TTC5. (**A**) Photo–cross-linking analysis of ^{35}S-labeled 94–amino acid RNCs of human β-tubulin and N-terminal mutants (indicated in red) with recombinant StrepII-tagged TTC5 containing the photo–cross-linking residue Bpa at position Phe^{194}. The nascent chain cross-link to TTC5 is indicated (TTC5-XL) and verified by pulldown via the StrepII tag (bottom panel). (**B**) Photo–cross-linking analysis using ^{35}S-labeled 64–amino acid RNCs of human β-tubulin or the N-terminal MRQI mutant. Wild-type or mutant recombinant StrepII-tagged TTC5 was included in the assay as indicated. The photo–cross-linking residue Bpa is at position 7 of the β-tubulin nascent chain. An aliquot of the total translation reaction was analyzed to verify equal levels of nascent chain (NC) synthesis by autoradiography and equal levels of recombinant TTC5 (SII-TTC5) by immunoblotting for the StrepII tag. The remainder was UV-irradiated, and TTC5 cross-links were recovered via the StrepII tag and visualized by autoradiography (bottom panel). (**C**) Summary of interaction analysis between the indicated recombinant TTC5 proteins and the indicated synthetic peptides in a thermal shift denaturation assay (see fig. S7). (**D**) Wild-type or mutant StrepII-tagged TTC5 was included during in vitro translation of wild-type or mutant 64–amino acid β-tubulin RNCs as indicated. Equal translation of ^{35}S-labeled nascent chain synthesis was verified (NC total). The remainder of each translation was affinity-purified via the StrepII tag and analyzed by staining of total proteins (bottom panel).

response to acute microtubule destabilization (Fig. 4A and fig. S8). Pulse labeling of wild-type cells with [^{35}S]methionine showed that of the major proteins visualized, tubulins were selectively reduced in their synthesis when cells are pretreated with microtubule-destabilizing agents (fig. S9). Selective reduction in tubulin protein synthesis was completely lost in TTC5 knockout cells, consistent with the failure to degrade tubulin mRNAs. Tubulin autoregulation, as judged by both mature mRNA versus pre-mRNA levels (Fig. 4A) and rates of protein synthesis (fig. S9), could be restored to TTC5 knockout cells by reexpression of wild-type TTC5 but not the peptide-binding mutant R147A or the ribosome-binding mutant KK-EE. Thus, TTC5 engagement of nascent tubulin at the ribosome is strictly required for tubulin mRNA degradation when cells initiate autoregulation. Access of TTC5 to the ribosome proved to be a regulated event.

The TTC5-RNC complex was found to be disrupted by a cytosolic factor whose activity was lost when cells were pretreated with colchicine to initiate autoregulation (Fig. 4B and fig. S10). Loss of this inhibitory activity under autoregulation conditions was accompanied by increased capacity of TTC5 to engage tubulin RNCs as measured by recovery of tubulin mRNAs (Fig. 4C and fig. S11). These results indicate that cells ordinarily contain an inhibitory factor that prevents TTC5 engagement of tubulin RNCs. This TTC5 inhibitor is inactivated when cells perceive excess αβ-tubulin, freeing TTC5 to engage tubulin RNCs and trigger tubulin mRNA degradation. TTC5's access to RNCs only during autoregulation explains why normally growing TTC5 knockout cells did not show notably elevated tubulin mRNA and protein (fig. S12). Because overexpressed TTC5 in the rescue cell lines did not trigger tubulin mRNA degradation until cells perceived excess αβ-tubulin (fig. S12), it seems that the inhibitor is not easily saturated. Further work will be necessary to identify the inhibitor and its mechanism of regulation.

Chromosome alignment and segregation during mitosis are sensitive to altered microtubule dynamics (*18*–*21*), motivating us to monitor these parameters in cells impaired in tubulin autoregulation (Fig. 4D). TTC5 knockout HeLa cells showed a higher rate of chromosome alignment errors in metaphase (by a factor of ~6.5) (Fig. 4E and fig. S13), a higher rate of segregation errors during anaphase (by a factor of ~2.4) (Fig. 4E and fig. S13), and a subtle but highly reproducible increase of mitotic duration (fig. S14). These phenotypes in TTC5 knockout cells were rescued by reexpression of wild-type TTC5 but not the peptide-binding or ribosome-binding mutants of TTC5 (Fig. 4E and figs. S13 and S14). Although the specific basis of mitotic defects in TTC5 knockout cells remains to be determined, we can ascribe the phenotypes to autoregulation, and not another TTC5 function (*22*–*24*), because the effects were not rescued by two unrelated point mutants of TTC5 that perturb autoregulation by different mechanisms.

TTC5 represents a highly selective and regulated ribosome-associating factor that only engages the ~2 to 3% of a cell's ribosomes that actively synthesize α- and β-tubulins. By marking tubulin-translating ribosomes, TTC5 is ideally positioned to recruit yet unidentified downstream effectors to this site that trigger mRNA decay. More generally, the translating ribosome represents a platform from which to effect abundance control of key cellular proteins because translation initiation (*25*), elongation (*26*), polypeptide fate (*27*), and mRNA stability (*28*) can all be locally regulated from this site. Specificity for particular substrates would be imparted by recognition of the nascent protein emerging from the ribosome exit tunnel. Thus, cells may contain a family of substrate-specific ribosome-associating factors analogous to TTC5 that dynamically tune the abundance of key proteins such as histones (*29*) and chaperones (*30*). The methods

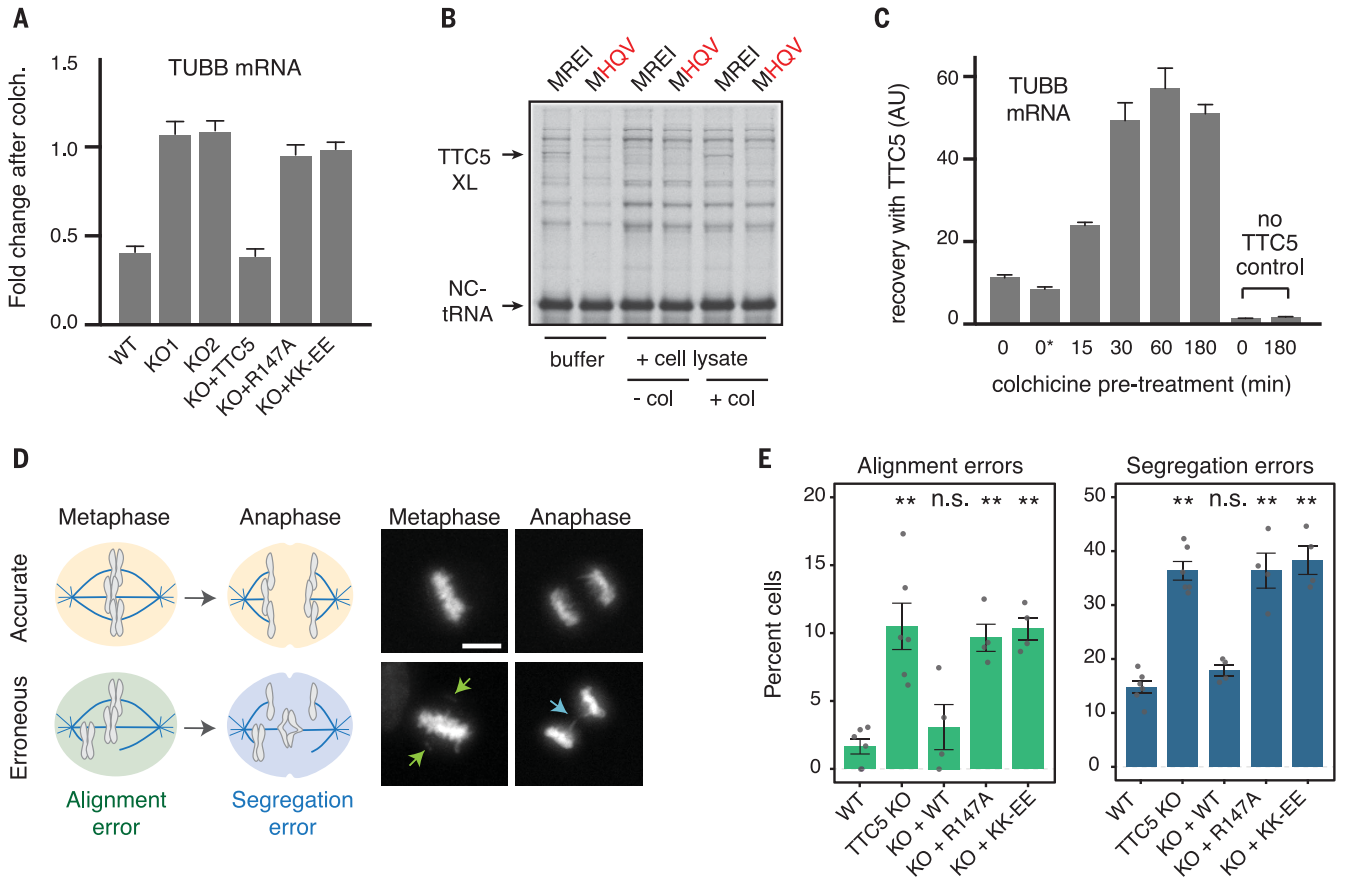


Fig. 4. TTC5 is required for tubulin autoregulation and accurate mitosis. (**A**) The indicated human embryonic kidney (HEK) 293 cell lines were either left untreated or treated for 3 hours with colchicine. The relative amounts of the indicated mRNAs or pre-mRNAs were quantified by reverse transcription quantitative polymerase chain reaction (RT-qPCR) and normalized to a control ribosomal RNA. Data are means ± SD from three replicates. Similar results are seen in HeLa cells and with different microtubule-destabilizing agents (fig. S8). (**B**) Preformed RNC-TTC5 complexes (see fig. S10) were mixed with buffer or cytosol from TTC5 knockout cells that had been pretreated (+col) or not (−col) with colchicine for 3 hours. All samples were subjected to UV cross-linking to monitor the nascent chain interactions. The positions of the non–cross-linked tRNA linked nascent chain (NC-tRNA) and TTC5 cross-link (TTC5-XL) are indicated. (**C**) TTC5-knockout HEK293 cells were pretreated for the indicated times with colchicine and used to prepare lysates. One of the control samples included colchicine added after cell lysis (indicated as 0*). The products recovered by binding to recombinant TTC5 were analyzed for β-tubulin mRNA by RT-qPCR. The relative recoveries are plotted (means ± SD from three replicates). Similar results were seen for α-tubulin and when nocodazole was used instead of colchicine to trigger autoregulation (fig. S11). (**D**) Diagram (left) and examples (right) of accurate (top) and erroneous (bottom) chromosome alignment and segregation visualized with SirDNA dye during mitosis in HeLa cell lines (see fig. S13). Scale bar, 5 μm. (**E**) Quantification of errors in chromosome alignment and segregation in the indicated HeLa cell lines. Data are means ± SEM from four to six independent biological replicates (dots) with 200 to 400 analyzed cells per replicate. $^{**}P < 0.001$ (paired Student t test); n.s., not significant.

and paradigm of cotranslational abundance control established here should provide a framework for studying this mode of cellular regulation.

REFERENCES AND NOTES

1. G. Borisy *et al.*, *Nat. Rev. Mol. Cell Biol.* **17**, 322–328 (2016).
2. A. Desai, T. J. Mitchison, *Annu. Rev. Cell Dev. Biol.* **13**, 83–117 (1997).
3. G. J. Brouhard, L. M. Rice, *Nat. Rev. Mol. Cell Biol.* **19**, 451–463 (2018).
4. R. A. Walker *et al.*, *J. Cell Biol.* **107**, 1437–1448 (1988).
5. I. Gasic, S. A. Boswell, T. J. Mitchison, *PLOS Biol.* **17**, e3000225 (2019).
6. D. W. Cleveland, M. A. Lopata, P. Sherline, M. W. Kirschner, *Cell* **25**, 537–546 (1981).
7. D. W. Cleveland, *Trends Biochem. Sci.* **13**, 339–343 (1988).
8. D. A. Gay, S. S. Sisodia, D. W. Cleveland, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 5763–5767 (1989).
9. J. S. Pachter, T. J. Yen, D. W. Cleveland, *Cell* **51**, 283–292 (1987).
10. T. J. Yen, P. S. Machlin, D. W. Cleveland, *Nature* **334**, 580–585 (1988).
11. C. J. Bachurski, N. G. Theodorakis, R. M. Coulson, D. W. Cleveland, *Mol. Cell. Biol.* **14**, 4076–4086 (1994).
12. N. G. Theodorakis, D. W. Cleveland, *Mol. Cell. Biol.* **12**, 791–799 (1992).
13. M. Gamerdinger *et al.*, *Mol. Cell* **75**, 996–1006.e8 (2019).
14. R. D. Wegrzyn *et al.*, *J. Biol. Chem.* **281**, 2847–2857 (2006).
15. B. Wiedmann, H. Sakai, T. A. Davis, M. Wiedmann, *Nature* **370**, 434–440 (1994).
16. M. Gamerdinger, M. A. Hanebuth, T. Frickey, E. Deuerling, *Science* **348**, 201–207 (2015).
17. UniProt Consortium, *Nucleic Acids Res.* **46**, 2699 (2018).
18. J. J. Vicente, L. Wordeman, *Curr. Opin. Cell Biol.* **60**, 36–43 (2019).
19. S. Petry, *Annu. Rev. Biochem.* **85**, 659–683 (2016).
20. S. L. Prosser, L. Pelletier, *Nat. Rev. Mol. Cell Biol.* **18**, 187–201 (2017).
21. S. L. Kline-Smith, C. E. Walczak, *Mol. Cell* **15**, 317–327 (2004).
22. J. T. Lynch, T. D. D. Somerville, G. J. Spencer, X. Huang, T. C. P. Somervaille, *Cell Death Dis.* **4**, e573 (2013).
23. C. Demonacos, M. Krstic-Demonacos, N. B. La Thangue, *Mol. Cell* **8**, 71–84 (2001).
24. X. Hu, R. D. Mullins, *J. Cell Biol.* **218**, 251–266 (2019).
25. N. Sonenberg, A. G. Hinnebusch, *Cell* **136**, 731–745 (2009).
26. J. C. Darnell, E. Klann, *Nat. Neurosci.* **16**, 1530–1536 (2013).
27. O. Brandman, R. S. Hegde, *Nat. Struct. Mol. Biol.* **23**, 7–15 (2016).
28. C. J. Shoemaker, R. Green, *Nat. Struct. Mol. Biol.* **19**, 594–601 (2012).
29. W. F. Marzluff, K. P. Koreski, *Trends Genet.* **33**, 745–759 (2017).
30. B. J. DiDomenico, G. E. Bugaisky, S. Lindquist, *Cell* **31**, 593–603 (1982).

ACKNOWLEDGMENTS

We thank V. Ramakrishnan for support and advice; S.-Y. Peak-Chew and M. Skehel for mass spectrometry analysis; J. Grimmett and T. Darling for advice, data storage, and high-performance computing; P. Emsley for advice; the MRC Laboratory of Molecular Biology EM Facility for microscopy support, sample preparation, and data collection; B. Raught and W. Harper for Flp-In TRex HeLa cells; and the Nikon Imaging Center at

Harvard Medical School for help with light microscopy. **Funding:** Supported by UK Medical Research Council grant MC_UP_A022_1007 (R.S.H.), NIH grant P50 107618 (T.J.M.), Harvard Medical School and Vallee Scholars Program (S.S.), and a Human Frontier Science Program postdoctoral fellowship (Z.L.). I.G. is a Merck Fellow of the Damon Runyon Cancer Research Foundation (DRG:2279-16). V.C. was supported by V. Ramakrishnan whose funding was from the MRC (MC_U105184332), the Wellcome Trust (WT096570), the Agouron Institute, and the Louis-Jeantet Foundation. **Author contributions:** Z.L. discovered TTC5 and performed all biochemical analyses; I.G. and S.S. generated and validated HeLa cell lines; I.G. designed and performed phenotypic analysis of mitosis; V.C. performed structural analysis of the TTC5-ribosome complex; N.P. set up and characterized the in vitro photo–cross-linking system; S.S., T.J.M., and R.S.H. supervised different aspects of the project; R.S.H. and Z.L. conceived the project, oversaw its implementation, and wrote the manuscript. All authors contributed to manuscript editing. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The cryo-EM map has been deposited to the EMDB (EMD-10380) and atomic coordinates have been deposited to the Protein Data Bank (PDB 6T59). All other data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6473/100/suppl/DC1
Materials and Methods
Figs. S1 to S15
Table S1
References (*31*–*49*)

View/request a protocol for this paper from *Bio-protocol*.

9 September 2019; accepted 31 October 2019
Published online 14 November 2019
10.1126/science.aaz4352

TOPOLOGICAL MATTER

Evidence for dispersing 1D Majorana channels in an iron-based superconductor

Zhenyu Wang[1,2], Jorge Olivares Rodriguez[1], Lin Jiao[1], Sean Howard[1], Martin Graham[3], G. D. Gu[4], Taylor L. Hughes[5], Dirk K. Morr[3], Vidya Madhavan[1]*

The possible realization of Majorana fermions as quasiparticle excitations in condensed-matter physics has created much excitement. Most studies have focused on Majorana bound states; however, propagating Majorana states with linear dispersion have also been predicted. Here, we report scanning tunneling spectroscopic measurements of crystalline domain walls (DWs) in $FeSe_{0.45}Te_{0.55}$. We located DWs across which the lattice structure shifts by half a unit cell. These DWs have a finite, flat density of states inside the superconducting gap, which is a hallmark of linearly dispersing modes in one dimension. This signature is absent in DWs in the related superconductor, FeSe, which is not in the topological phase. Our combined data are consistent with the observation of dispersing Majorana states at a π-phase shift DW in a proximitized topological material.

Majorana fermions are putative elementary particles that are their own antiparticles (*1*). Emergent analogs of these fermions have been argued to exist as quasiparticle excitations in condensed-matter systems (*2*–*7*) and have attracted much attention as possible building blocks of fault-tolerant quantum computation (*8*, *9*). So far, various predictions and realizations of localized Majorana bound states (MBS) have been reported. The platforms include strong spin-orbit–coupled semiconductor nanowires (*10*–*14*), ferromagnetic atomic chains (*15*–*17*), and topological insulators that are proximity-coupled with s-wave superconductors (*18*, *19*); in all of these cases, the MBS were spectroscopically identified as zero-energy conductance anomalies. In addition to the localized MBS, however, theoretical predictions show that dispersing Majorana states may also be realized as quasiparticles in condensed-matter systems (*18*, *20*–*22*). These quasiparticles are of fundamental interest and may be harnessed for quantum computing. Dispersive Majorana modes have been proposed as edge states of chiral p-wave superconductors, in hybrid systems that combine superconductors with a quantum anomalous Hall insulator (*23*), or two-dimensional (2D) magnetic Fe islands (*24*, *25*). However, these platforms are difficult to fabricate; moreover, most of them are stable only at low temperatures. This makes future applications highly challenging in these systems.

Iron-based superconductors provide an alternative pathway for pursuing Majorana modes at higher temperatures. $FeSe_xTe_{1-x}$ [Fe(Se,Te)] is the simplest compound in the Fe-based superconductor family, with an optimum critical temperature (T_c) of 14.5 K. This family of materials is highly attractive owing to its versatility and tunability. The materials grow well in thin film form (*26*), and their T_c can be substantially enhanced through doping, pressure, and strain (*27*). Through density functional theory, it has been found that for a range of concentrations around 50% Se, Fe(Se,Te) possesses helical Dirac surface states owing to band inversion along the Γ-Z direction (*28*–*31*). In accordance with the Fu and Kane model (*18*), when an s-wave superconducting gap opens in the Dirac surface states (because of proximity to s-wave superconductivity in the bulk), it provides the ideal conditions for hosting MBS. There are multiple pieces of supportive evidence for this scenario in Fe(Se,Te): High-resolution angle-resolved photoemission spectroscopy (ARPES) data reveal helical surface states that exhibit an s-wave gap below T_c, and a sharp zero-bias peak has already been observed inside vortex cores (*32*–*34*).

s-Wave–proximitized topological surface states can also host time-reversed pairs of dispersing 1D Majorana states along domain walls (DWs) separating regions in which the superconducting order parameter is phase-shifted by π (*18*). These modes possess a linear dispersion ($E = \pm vk_y$) with momentum parallel to the DW. This linear dispersion in one dimension implies a constant density of states (DOS) for energies below the superconducting gap—one of the key experimental signatures of dispersing Majorana states.

In this work, we used scanning tunneling microscopy (STM) to interrogate crystalline DWs in the proximitized Dirac surface states of $FeSe_{0.45}Te_{0.55}$ in a search for signatures of 1D dispersing Majorana modes.

As with most iron-based superconductors, the Fermi surface of $FeSe_{0.45}Te_{0.55}$ is composed of two hole pockets (Fig. 1A, α′ in red and β in green) around the Γ-point and two electron pockets (γ in blue) at the Brillouin zone corner (M-point). According to theory, Te substitution into FeSe shifts the bulk p_z band [found above the Fermi energy (E_F) in FeSe] downward toward the Fermi level (*29*). This band then hybridizes with the d_{xz} band (α band) to create a topological band inversion that pushes the α band ~14 meV below E_F. In the resulting band gap, a topological Dirac surface state emerges, centered at the Γ-point on the (001) surface (Fig. 1B). Below T_c, superconducting gaps are expected to open on both the surface and bulk bands (*32*).

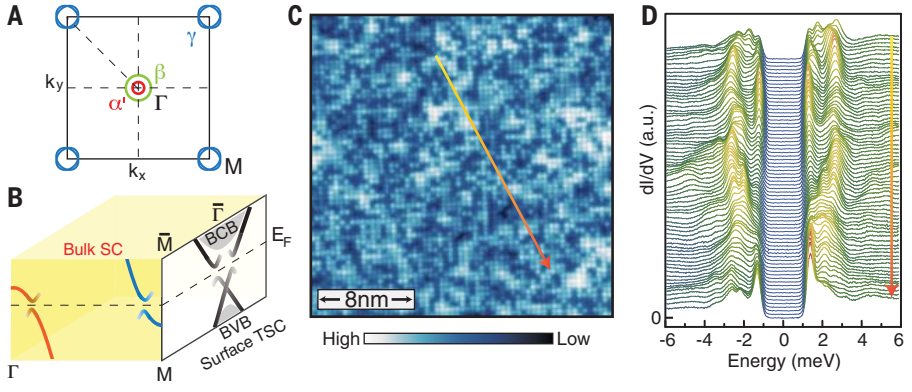


Fig. 1. Band structure and superconductivity in $FeSe_{0.45}Te_{0.55}$. (**A**) Sketch of bulk Fermi surfaces of Fe(Se,Te) at momentum $k_z = 0$. (**B**) Cartoon image showing superconductivity in the bulk and proximitized superconductivity in the topological surface state (*31*). (**C**) Topographic image in a 25- by 25-nm field of view (bias voltage V_S = 40 mV, tunneling current I_t = 100 pA). (**D**) Scanning tunneling spectroscopy (STS) data taken along the line shown in (C) at 0.3 K [V_S = 6 mV, I_t = 300 pA, modulation voltage (V_{mod}) = 58 μV]. The spectra are vertically offset for clarity.

[1]Department of Physics and Frederick Seitz Materials Research Laboratory, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA. [2]Department of Physics, University of Science and Technology of China, Hefei, Anhui 230026, China. [3]Department of Physics, University of Illinois at Chicago, Chicago, IL 60607, USA. [4]Condensed Matter Physics and Materials Science Department, Brookhaven National Laboratory, Upton, NY 11973, USA. [5]Department of Physics and Institute for Condensed Matter Theory, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.
*Corresponding author. Email: vm1@illinois.edu

In Fig. 1C, we show an atomically resolved STM topographic image of the chalcogen (Se/Te) surface layer, where the Te atoms appear brighter than Se because of their more extended electronic orbitals. There is a marked absence of interstitial Fe atoms on the surface, usually observed as bright protrusions in the topography (*35*, *36*). This, in combination with the sharp superconducting transition (fig. S1) (*37*), confirms the high quality of these samples. We measured a series of differential conductance (dI/dV) spectra along the line shown in Fig. 1C and present them in Fig. 1D. The data show that the spectral weight is completely suppressed to zero over a finite energy range ~±1 meV, and sharp peaks appear near the gap edge. These observations strongly suggest that there is no nodal structure in the gap function of $FeSe_{0.45}Te_{0.55}$, and the gap minima, if anisotropy exists, should be larger than 1 meV.

There is an ongoing controversy regarding the gap values for each band reported by various ARPES and optical conductivity studies on similar materials (*38*–*41*). STM is the ideal probe with which to measure gap values on different bands with high accuracy. However, because of the inhomogeneity caused by doping in Fe(Se,Te), the gap values and thus the number of gaps seen in any one spectrum show spatial variations (Fig. 1D). To obtain statistical information on the gap values and distribution, we recorded tunneling spectra $dI(r, V)/dV$ on a densely spaced grid (240 by 240) over a 100- by 100-nm field of view (FOV). The gap values were extracted through our multigap-finding algorithm, which finds the position of peak and shoulder features in each dI/dV spectrum and accepts them as coherence peaks if they are particle-hole symmetric (*37*). We classify the results by the number of gaps found for each spectrum and show a color-coded 2D map (gap map) in Fig. 2A.

We found spectra with one, two, or three gaps in the energy range (–3.5 meV, 3.5 meV). To visualize the evolution of the spectra as a function of position, a spectral line cut traversing the three regions (white line on the gap map) is shown in Fig. 2B. One can see the spectra evolving from displaying two gaps to three and then to a single gap. The statistical analysis of the gap magnitudes divided by category (colored histograms), as well as the overall results, are shown in Fig. 2C. For ~20% of the spectra taken in this FOV, we can distinguish only one gap centered around 1.4 meV; the two-gap spectra cover ~57% of the FOV area, with mean gap values around 1.4 and 2.4 meV; in the remaining area, three gaps can be detected simultaneously, with mean values of 1.4, 1.9, and 2.4 meV. These data suggest that multiple gaps exist at all points of the sample, but statistical variations in their magnitude sometimes prevent us from resolving

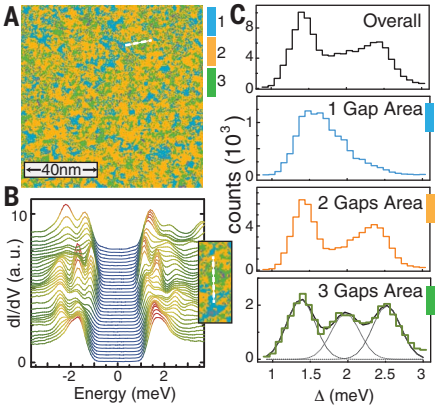


Fig. 2. Statistical analysis of superconducting gaps. (**A**) 100- by 100-nm map depicting the distribution of superconducting gaps in $FeSe_{0.45}Te_{0.55}$. Blue, orange, and green colors indicate whether a single gap, two gaps, or three gaps were found at each pixel, respectively. The gap values at each pixel were obtained through a multipeak-finding algorithm (37). (**B**) STS spectra at 0.3 K obtained along the white line shown in (A), starting (top to bottom) with a two-gap region (orange), which transitions to a three-gap region (green) and ends in a single-gap region (blue). The spectra are vertically offset for clarity. (**C**) Histogram of the gap values in the one-, two-, and three-gap regions. The dark curves show Gaussian fits to the gap distribution, with mean gap values of 1.4, 1.9, and 2.4 meV.

them individually. The largest gap value extracted from ARPES is around 4 to 5 meV (*40*). This larger gap is only seen as a hump-like feature in our spectra and was not picked up in our gap map because of the suppressed intensity of the corresponding peaks. However, this feature can be seen in line cuts (fig. S2) (*37*). If we assign this hump-like feature (around ±4.5 meV) to be the coherence peaks arising from the large superconducting gap on the γ sheet as shown in (*40*), then, by comparison with ARPES data, the mean gap values of 1.4 and 2.4 meV may be assigned to the smaller gaps on the α′ and β bands, respectively. This suggests that consistent with recent ARPES data, the 1.9-meV gap may be assigned to the topological surface state, indicating the topological nature of these samples (*32*).

Recent STM measurements in Fe(Se,Te) have reported the existence of zero-bias conductance peaks inside vortex cores and near interstitial Fe atoms (*33*, *34*, *36*), which have been proposed to be signatures of zero-dimensional MBS. We observed similar spectral line shapes inside several vortex cores and near–atomic-scale defects (figs. S3 and S4) in our samples, all of which are consistent with a topologically nontrivial surface state (*37*). Here, we report the existence of 1D dispersing Majorana modes near a domain-wall defect. This defect was discovered with atomically resolved topography as a 1D feature on the surface represented by a bright line (Fig. 3A). A zoomed-in view (Fig. 3D) reveals that this bright line separates two crystal domains where the lattice shows a relative phase shift. This shift is reflected as a split in reciprocal-space Bragg peaks of the Fourier transform of the image (Fig. 3B and fig. S5) (*37*). The magnitude of the split in reciprocal space corresponds to a spatial scale of 12 nm (half of the FOV of Fig. 3A), which is consistent with the domain size in this FOV. To determine the magnitude of the lattice shift between the two sides of the DW, we carried out a displacement analysis (fig. S6), which maps the relative phase of the lattice on either side (*37*). The analysis indicates that the lattice undergoes a half-unit-cell shift across such a DW. This half-unit-cell shift, as we will show later, is essential for the realization of dispersing Majorana modes.

Differential conductance spectra obtained along three distinct paths traversing the DW (Fig. 3F) reveal an intriguing evolution. As one approaches the DW, the superconducting coherence peaks in the dI/dV spectra are suppressed, and new electronic states emerge inside the gap, resulting in a nearly featureless, constant finite dI/dV inside the gap at the DW center as indicated with the highlighted lines in Fig. 3F and the inset (fig. S7) (*37*). Despite the constant DOS inside the superconducting gap, the DOS at the DW still exhibits superconducting signatures (Fig. 3F, inset, and fig. S8H) , which indicates that the flat DOS is not simply caused by a region of normal metal (*37*). One explanation for this observation is the presence of linearly dispersing Majorana states at the DW because it would naturally give rise to a constant dI/dV in one dimension.

According to the Fu-Kane model (*18*), realizing 1D dispersing Majorana states requires three ingredients: nontrivial topological surface states, s-wave superconductivity that gaps the surface states, and a π-phase shift in the superconducting order parameter across the DW. Our detailed gap maps already indicate the presence of proximity-gapped Dirac surface states, thus satisfying the first two criteria. This leaves us with the question of how to generate a superconducting phase shift. For the pairing symmetries allowed in Fe(Se,Te), it is possible to have an interplay between the crystal structure and the phase of the superconducting order parameter. One possibility comes from the predicted odd parity s-wave pairing in iron-based superconductors, which encodes pairing between next-nearest-neighbor sites (*42*, *43*) in the 2-Fe unit cell. In this case, the order parameters on the two Fe-sublattices have a π-phase difference. A half-unit-cell shift of the lattice in such a system would naturally create a π-phase shift across the domains (Fig. 3E). It has also been argued that the s± pairing in iron-based superconductors can

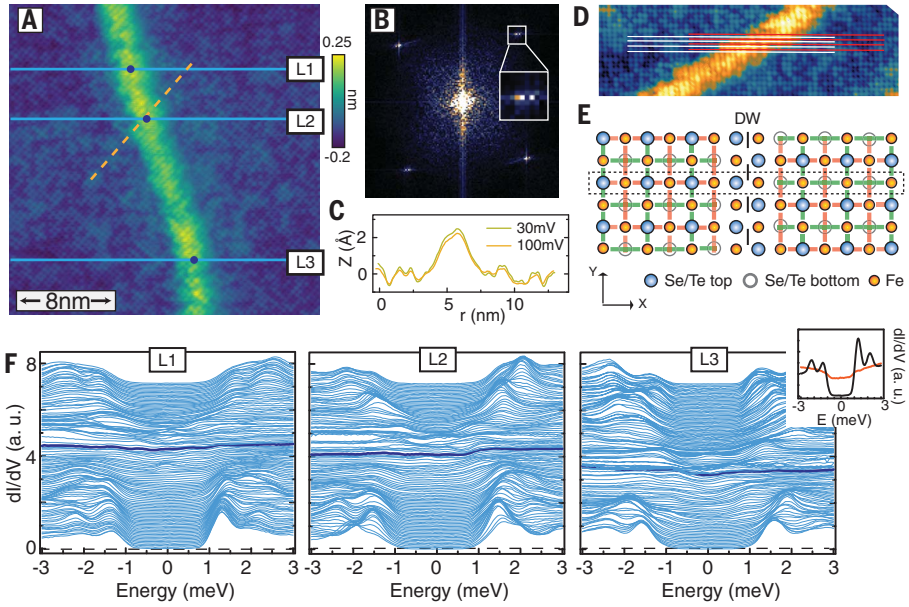


Fig. 3. Signature of dispersing 1D Majorana modes at a DW. (**A**) 25- by 25-nm topographic image showing a DW (bright line) (V_S = 4 mV, I_t = 250 pA). (**B**) 2D fast Fourier transform of (A), showing a splitting of the Bragg peaks, which indicates the presence of domains in this image. (Inset) Zoom-in near one of the Bragg peaks. (**C**) Height scans taken at different bias voltages along the yellow dashed line in (A). (**D**) Zoom-in of the DW. The white and red lines track the atomic lattice on both sides of the DW. A half-unit-cell shift can be observed between one side and the other. (**E**) Schematic of the half-unit-cell shift across the DW. The schematic also depicts how one might obtain a π-phase shift in the superconducting order parameter across such a DW. Superimposed on the lattice are red and green bars, which denote the parity of next-nearest-neighbor pairing (*42*, *43*). As an example, tracking the atoms inside the dashed box, one can see that the parity shifts from red on the left of the DW to green on the right. This creates a π-phase shift in the superconducting order parameter. (**F**) Line-cut profiles of *dI/dV* spectra along the three blue lines in (A), which cross the DW (V_S = 4 mV, I_t = 250 pA, V_{mod} = 58 μV). The spectra shapes obtained right on the DW [at position of dots in (A)] are highlighted with a dark blue color. For clarity, the spectra are vertically offset from each other by 0.06 (6 nS). (Inset) A direct comparison of the spectra taken on the DW (orange) and far away (black). All data were obtained at 0.3 K.

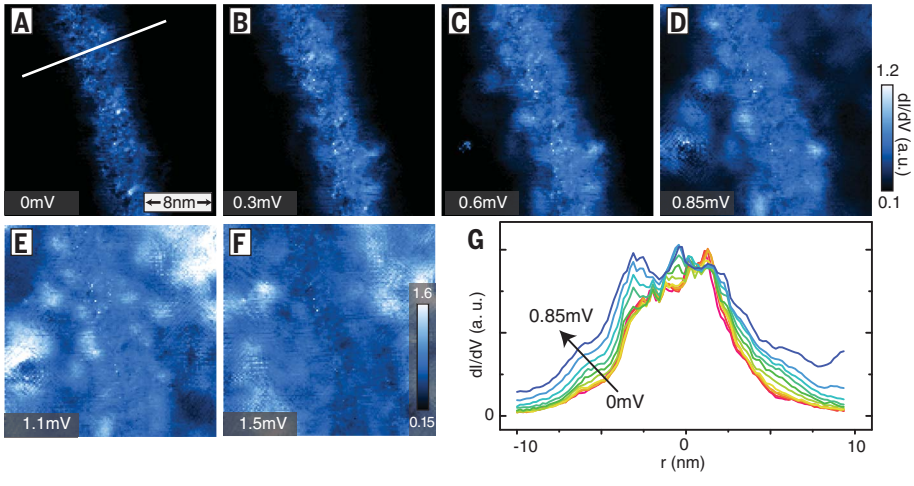


Fig. 4. Spatial distribution of the 1D Majorana mode at a DW with increasing energy. (**A** to **F**) *dI/dV* maps from 0 to 1.5 meV at 0.3 K. The maps are 25 by 25 nm in size, and spectra were obtained on a 130- by 130-pixel grid. DW states are present at all energies inside the gap up to 1 meV, when the states merge into the coherence peaks. However, the spatial extent of the states grows with increasing energy. (**G**) DOS profiles measured at different energies along the white line perpendicular to the DW.

generate a π-phase shift at the intersections of crystal terminations with different orientations (*44*). Thus, Fe(Se,Te) is an excellent candidate, with all the essential ingredients necessary for hosting dispersing Majorana modes.

To distinguish this possibility from other scenarios, we studied other extended defects (figs. S9 and S10) that are not expected to give rise to a π-phase shift and found that the flat DOS signature is absent (*37*); 1D defects without the half-unit-cell shift have the effect of decreasing the gap magnitude (fig. S9), whereas step edges, which are strong potential scatterers, induce resonant bound states inside the superconducting gap (fig. S10). Furthermore, the half-unit-cell shift DW is spectroscopically similar to the rest of the sample above T_c. As shown in fig. S8, at 14 K the spectra both on and away from the DW are almost identical (*37*). This indicates that the DW does not have a noticeable effect on the local electronic structure above T_c, and its effects become prominent only below T_c.

One might wonder whether the experimentally observed DW modes could also possess a topologically trivial origin, unrelated to existence of a topological surface state. On the basis of previous studies, the superconducting order parameter in Fe(Se,Te) is expected to be a sign-changing s± state (*45*, *46*). In such a state, defects, regardless of their magnetic properties, would induce impurity states inside the superconducting gap. The experimentally observed DW representing a 1D defect could therefore lead to the emergence of an impurity band inside the superconducting gap even in a topologically trivial phase. To investigate this possibility, we used a theoretical model for a topologically trivial superconducting state of Fe(Se,Te) (*47*) and represented the DW as a line of potential scatterers (fig. S11) (*37*). We found, as expected, that the DW gives rise to impurity states inside the superconducting gap. However, these states do not in general traverse the superconducting gap (only for fine-tuned values of the scattering potential do impurity states near zero energy emerge). Moreover, such states are not uniformly distributed in energy inside the gap and cannot result in the observed constant density of states. The same conclusion holds if the DW-separated, π-phase–shifted superconducting regions are present in an otherwise topologically trivial phase (fig. S12) (*37*). These findings are further confirmed by our experimental study of twin DWs in the topologically trivial but related superconducting compound, FeSe (fig. S13) (*37*). Although such DWs give rise to a suppression of the superconducting gap, they do not result in a constant DOS. These theoretical and experimental findings taken together make it unlikely that the observed constant DOS near the DW can arise in a topologically trivial superconducting phase, which

further emphasizes the important role played by nontrivial topology in creating the observed DW modes.

The topological nature of the DW modes necessarily dictates a specific spatial and energy dependence of the DW states. In particular, with increasing energy, the Majorana mode must continuously evolve from being localized at the DW at zero energy to being delocalized as their energy reaches the gap edge. In other words, the modes' localization length increases with increasing energy, leading to an increase in the modes' spatial extent from the DW. To visualize this evolution, we obtained spatial differential conductance maps (dI/dV maps) in the vicinity of the DW ranging in energies from E_F up to 1.5 mV, where the first set of coherent peaks is located (Fig. 4, A to F). At E_F, the in-gap states are confined within an ~3-nm width. These states begin to expand in real space with increasing energy and become far less visible at 0.85 mV because of a lack of contrast in the intensity with respect to the rest of the area. The dI/dV maps show that the DW states are present at all energies inside the gap. However, the spatial extent of the states grows with increasing energy, which is consistent with a topological origin for these modes.

$FeSe_{0.45}Te_{0.55}$ may have provided the first glimpse into linearly dispersing 1D Majorana modes. Possible future experiments include the measurement of the fractional (4π) Josephson effect using a flux loop or depositing magnetic layers to generate chiral Majorana modes or localized MBS. Beyond the Majorana fermion context, our experimental results have two important implications. First, our observations provide supporting evidence for the existence of topological surface states and a Fu-Kane proximitized superconducting state in $FeSe_{0.45}Te_{0.55}$. Second, the connection between the crystal DW and the superconducting π-phase shift provides evidence in support of a superconducting order parameter with a real-space sign change within a unit cell (*42*).

REFERENCES AND NOTES

1. E. Majorana, *Nuovo Cim.* **14**, 171–184 (1937).
2. N. Read, D. Green, *Phys. Rev. B* **61**, 10267–10297 (2000).
3. A. Kitaev, *Ann. Phys.* **321**, 2–111 (2006).
4. D. A. Ivanov, *Phys. Rev. Lett.* **86**, 268–271 (2001).
5. J. Alicea, *Rep. Prog. Phys.* **75**, 076501 (2012).
6. C. W. J. Beenakker, *Annu. Rev. Condens. Matter Phys.* **4**, 113–136 (2013).
7. S. R. Elliott, M. Franz, *Rev. Mod. Phys.* **87**, 137–163 (2015).
8. A. Kitaev, *Ann. Phys.* **303**, 2–30 (2003).
9. C. Nayak, S. H. Simon, A. Stern, M. Freedman, S. Das Sarma, *Rev. Mod. Phys.* **80**, 1083–1159 (2008).
10. R. M. Lutchyn, J. D. Sau, S. Das Sarma, *Phys. Rev. Lett.* **105**, 077001 (2010).
11. Y. Oreg, G. Refael, F. von Oppen, *Phys. Rev. Lett.* **105**, 177002 (2010).
12. V. Mourik *et al.*, *Science* **336**, 1003–1007 (2012).
13. A. Das *et al.*, *Nat. Phys.* **8**, 887–895 (2012).
14. S. M. Albrecht *et al.*, *Nature* **531**, 206–209 (2016).
15. S. Nadj-Perge *et al.*, *Science* **346**, 602–607 (2014).
16. M. Ruby *et al.*, *Phys. Rev. Lett.* **115**, 197204 (2015).
17. H. Kim *et al.*, *Sci. Adv.* **4**, eaar5251 (2018).
18. L. Fu, C. L. Kane, *Phys. Rev. Lett.* **100**, 096407 (2010).
19. H. H. Sun *et al.*, *Phys. Rev. Lett.* **116**, 257003 (2016).
20. A. P. Schnyder, S. Ryu, A. Furusaki, A. W. W. Ludwig, *Phys. Rev. B* **78**, 195125 (2008).
21. X. L. Qi, T. L. Hughes, S. Raghu, S.-C. Zhang, *Phys. Rev. Lett.* **102**, 187001 (2009).
22. Y. Tanaka, M. Sato, N. Nagaosa, *J. Phys. Soc. Jpn.* **81**, 011013 (2012).
23. Q. L. He *et al.*, *Science* **357**, 294–299 (2017).
24. S. Rachel, E. Mascot, S. Cocklin, M. Vojta, D. K. Morr, *Phys. Rev. B* **96**, 205131 (2017).
25. A. Palacio-Morales *et al.*, *Sci. Adv.* **5**, eaav6600 (2019).
26. P. Mele, *Sci. Technol. Adv. Mater.* **13**, 054301 (2012).
27. K. Horigane, N. Takeshita, C.-H. Lee, H. Hiraka, K. Yamada, *J. Phys. Soc. Jpn.* **78**, 063705 (2010).
28. N. Hao, J. Hu, *Phys. Rev. X* **4**, 031053 (2014).
29. Z. Wang *et al.*, *Phys. Rev. B* **92**, 115119 (2015).
30. X. Wu, S. Qin, Y. Liang, H. Fan, J. Hu, *Phys. Rev. B* **93**, 115129 (2016).
31. G. Xu, B. Lian, P. Tang, X.-L. Qi, S.-C. Zhang, *Phys. Rev. Lett.* **117**, 047001 (2016).
32. P. Zhang *et al.*, *Science* **360**, 182–186 (2018).
33. D. Wang *et al.*, *Science* **362**, 333–335 (2018).
34. T. Machida *et al.*, *Nat. Mater.* **18**, 811–815 (2019).
35. M. Chen *et al.*, *Nat. Commun.* **9**, 970 (2018).
36. J. Yin *et al.*, *Nat. Phys.* **11**, 543–546 (2015).
37. Materials and methods are available as supplementary materials.
38. B. Zeng *et al.*, *Nat. Commun.* **1**, 112 (2010).
39. K. Okazaki *et al.*, *Phys. Rev. Lett.* **109**, 237011 (2012).
40. H. Miao *et al.*, *Phys. Rev. B* **85**, 094506 (2012).
41. C. C. Homes *et al.*, *J. Phys. Chem. Solids* **72**, 505–510 (2011).
42. J. Hu, *Phys. Rev. X* **3**, 031004 (2013).
43. N. Hao, J. Hu, *Phys. Rev. B* **89**, 045144 (2014).
44. R. X. Zhang, W. S. Cole, S. Das Sarma, *Phys. Rev. Lett.* **122**, 187001 (2019).
45. T. Hanaguri, S. Niitaka, K. Kuroki, H. Takagi, *Science* **328**, 474–476 (2010).
46. P. J. Hirschfeld, M. M. Korshunov, I. Mazin, *Rep. Prog. Phys.* **74**, 124508 (2011).
47. S. Sarkar *et al.*, *Phys. Rev. B* **96**, 060504 (2018).
48. Z. Wang *et al.*, Replication data for: Evidence for dispersing 1D Majorana channels in an iron-based superconductor. Zenodo (2019); 10.5281/zenodo.3529111

ACKNOWLEDGMENTS

The authors thank H. Ding, S. Kivelson, and C. Kane for useful conversations. **Funding:** The work in Brookhaven is supported by the Office of Science, U.S. Department of Energy, under contract DE-SC0012704. V.M. gratefully acknowledges support from U.S. Department of Energy, Office of Science, Basic Energy Sciences, under award DE-SC0014335 (STM studies) and NSF award DMR-1610143. T.L.H. thanks the U.S. National Science Foundation under the Materials Research Science and Engineering Center program under NSF award DMR-1720633 (SuperSEED) for support. M.G. and D.K.M. acknowledge support from the U.S. Department of Energy, Office of Science, Basic Energy Sciences, under award DE-FG02-05ER46225. **Author contributions:** V.M. and Z.W. conceived the experiments; Z.W., J.O.R., and L.J. performed the experiments; J.O.R. and S.H. wrote software for data analysis; Z.W., J.O.R., and S.H. carried out data analysis; G.D.G. made the samples; M.G., T.L.H., and D.K.M. carried out the theory; and V.M. and Z.W. wrote the paper, with contributions from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The experimental data and theory code of this study are available at (*48*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6473/104/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S13
Tables S1 and S2
References (*49*, *50*)

29 January 2019; resubmitted 20 July 2019
Accepted 14 November 2019
10.1126/science.aaw8419

High-Definition MWIR Cameras
Compact in design, FLIR Systems A8200sc and A8300sc midwave infrared (MWIR) cameras offer the speed and sensitivity researchers need to capture rapid thermal events and record the most accurate temperature measurements. Incorporating cooled indium antimonide detectors, they produce 1,280 pixel × 1,280 pixel (A8200sc) or 1,280 pixel × 720 pixel (A8300sc) thermal images with very low noise. The cameras capture the finest image details, making them perfect for electronics inspection, aerial thermal mapping, nondestructive material testing, and industrial R&D applications. Operating from 3 µm to 5 µm, they can stream crisp, low-noise HD data to a PC for live-image viewing and recording, and also provide full user control over integration times, synchronization inputs, and window sizes.
FLIR Systems
For info: +32-(0)-3665-5100
www.flir.co.uk/products/a8300sc

Antibiotic-Resistant Bacteria Molecular Screening Test
The BD Check-Points CPO Assay can detect the five most-common carbapenemase genes in less than 2.5 h, and runs on BD's fully automated molecular platform, the BD MAX. The new carbapenemase-producing organisms (CPO) detection assay is an addition to BD's comprehensive portfolio dedicated to reducing transmission and outbreak of deadly pathogens. It allows hospitals and clinical laboratories to offer a quick and efficient time to result, enabling immediate deployment of control measures and treatment.
Becton, Dickinson and Company
For info: 201-847-6800
www.bd.com

Western Blot Blocking Reagent
Bio-Rad Laboratories announces EveryBlot Blocking Buffer, a Western blot blocking reagent that offers 5-min blocking time and produces higher signal and lower background levels—resulting in greater sensitivity for Western blots compared to other available blocking buffers. It features superior sensitivity and low background with both chemiluminescent and fluorescence detection across a wide range of targets. EveryBlot is compatible with all detection methods and also works with detection of phosphorylated targets, which eliminates the need for optimization.
Bio-Rad Laboratories
For info: 800-424-6723
www.bio-rad.com/seeeveryblot

Basal Medium and Feed
Lonza now offers eCHO Basal Medium and Feed, designed to optimize protein production by delivering a consistent yield and highly viable cells while keeping lactate levels to a minimum for optimal protein integrity. Researchers working in the manufacture of recombinant proteins for medical use can now benefit from this new serum-free, chemically defined, hydrolysate-free, and nonanimal origin medium to advance cell proliferation and enhance productivity. Chinese hamster ovary (CHO) cells are commonly used in the manufacture of biotherapeutics, such as monoclonal antibodies, complex glycoproteins, and biosimilars. Combining basal and feed medium into a one-powder product, eCHO Medium enhances late-stage cell viability, accelerating protein purification and allowing for production of a consistently increased amount of proteins per cell. eCHO is available in both liquid and powder formats to suit varying application-throughput needs.
Lonza
For info: 800-638-8174
www.lonza.com/echo

Cancer and Normal Tissue cDNA Arrays
TissueScan Cancer and Normal Tissue cDNA Arrays are developed for differential gene expression analysis and validation among hundreds of different human tissues in less than 2 h. Tissue complementary DNA (cDNA) of each array is synthesized from high-quality total RNA of pathologist-verified tissues, normalized and validated with ß-actin in two sequential quantitative PCR (qPCR) analyses, and provided with clinical information and quality control data. All TissueScan cDNA Arrays are in PCR microplates from Abgene (AB-0600). According to the plate manufacturer, the microplates will fit in standard thermocyclers and qPCR thermocyclers. Over 20 different cancer tissues are available, including breast, colon, prostate, melanoma, lymphoma, and ovary.
OriGene Technologies
For info: 888-267-4436
www.origene.com

Live-Cell Imaging System
Our MuviCyte live-cell imaging system is designed to operate inside your cell culture incubator, so you can keep your cells healthy and perform assays over hours, days, or even weeks. With the MuviCyte system, you can run a variety of assays with three-color fluorescence imaging, z-stacking, and stitching; use your preferred culture vessels—chamber slides, Petri dishes, T-flasks, or microplates; and image automatically for walkaway convenience. Put that together with MuviCyte's flexible moviemaking software, and you have a great way to gain more realistic and meaningful insights from your cells.
PerkinElmer
For info: 800-762-4000
www.perkinelmer.com

Micropumps
TTP Ventus, a manufacturer of high-performance, silent micropumps for medical and life science markets, has announced the introduction of a range of new pumps that extend performance, delivering high flow rates, high pressures, and long lifetimes. This opens up new product innovation possibilities for applications including microfluidics research, medical diagnostics, therapeutic devices, and wearables. The new HP Series has been designed for applications requiring high pressure, with the first model delivering in excess of 800 mBar, doubling the capability of previous models. The new LT Series was developed for applications requiring long lifetimes, typically 5,000 h or more. TTP Ventus has also extended its BL and XP Series with four additional pumps. These provide increased flow rates of up to 2.1 L/min, enabling faster inflation with blood pressure cuffs or therapeutic cushions, for example.
TTP Ventus
For info: +44-(0)-1763-262626
www.ttpventus.com/products

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

UNIVERSITY OF MINNESOTA
Medical School

Bacaner Endowed Chair in Physiology

The department of Integrative Biology and Physiology at the University of Minnesota Medical School is recruiting a tenured faculty member for the Bacaner Endowed Chair in Physiology. We seek exceptional faculty candidates with an established research program in integrative systems physiology, spanning from gene to whole organism, with specific interests in the cardiovascular system, metabolism (diabetes/obesity), muscle physiology, stress physiology and related fields. Substantial resources, with state-of-the-art research space, is committed to this effort. Successful candidates will have an established and innovative research program that embraces biological complexity and physiology from molecular building blocks to the living organism and human disease. Applicants must have a strong record of research accomplishments, as documented by publications in leading peer-reviewed journals along with an outstanding record of extramural funding. Minimum requirements are a PhD, MD or MD/PhD with tenure at an academic research institution. The Department is a central part of the newly established Biomedical Discovery District at the UMN and hosts a state-of-the-art Core facility for cardio-metabolic phenotyping (IBP Phenotyping Core). The Department is an integral component of Institutes/Centers including the Institute for Diabetes, Obesity and Metabolism, Lillehei Heart Institute, Institute for the Biology of Aging and Metabolism, Masonic Cancer Center, Center for Immunology, with exceptional on-site research core facilities. There are also several active and relevant training grant (T32) programs that offer support for training students and fellows. Additional information about the department can be found at https://www.physiology.umn.edu.

To apply, please upload a letter of interest, curriculum vitae, statement of programmatic and research goals in response to job opening ID 334734 at **http://www1.umn.edu/ohr/employment**. Please also arrange to have three (3) letters of recommendation sent to jotto@umn.edu or Search Committee, Department of Integrative Biology and Physiology, Medical School, 6-125 Jackson Hall, 321 Church Street S.E., Minneapolis, MN 55455. Review of applications will begin **February 3, 2020** and continue until the position is filled.

The University of Minnesota provides equal access to and opportunity in its programs, facilities, and employment without regard to race, color, creed, religion, national origin, gender, age, marital status, disability, public assistance status, veteran status, sexual orientation, gender identity, or gender expression. To learn more about diversity at the U: http://diversity.umn.edu.

THE UNIVERSITY OF CHICAGO

Academic Gastroenterologist – Hepatology
Biological Sciences Division: Department of Medicine

Position Description

The University of Chicago's Department of Medicine, Section of Gastroenterology, Hepatology and Nutrition, is searching for a full-time faculty member at any rank to join our rapidly growing Center for Liver Diseases. The appointee will have a mix of outpatient and inpatient clinical responsibilities, including general hepatology and liver transplant service and consultation at multiple sites in the Chicago area. Other duties will include teaching and supervision of trainees and students, and scholarly activity. We especially welcome applicants with advanced training in transplant hepatology. Academic rank and compensation (including a generous package of fringe benefits) are dependent upon qualifications.

Prior to the start of employment, qualified applicants must: (1) have an medical doctorate or equivalent, (2) hold or be eligible for medical licensure in the State of Illinois, (3) be Board Certified or Eligible in Gastroenterology, and (4) have expertise in clinical and research (basic or translational) aspects of nonalcoholic fatty liver disease (NAFLD).

To be considered, those interested must apply through The University of Chicago's Academic Recruitment job board, which uses Interfolio to accept applications: **http://apply.interfolio.com/72324**. Applicants must upload: a CV including bibliography and cover letter. Review of applications ends when the position is filled.

Equal Employment Opportunity Statement

We seek a diverse pool of applicants who wish to join an academic community that places the highest value on rigorous inquiry and encourages diverse perspectives, experiences, groups of individuals, and ideas to inform and stimulate intellectual challenge, engagement, and exchange.

The University of Chicago is an Affirmative Action/Equal Opportunity/Disabled/Veterans Employer and does not discriminate on the basis of race, color, religion, sex, sexual orientation, gender identity, national or ethnic origin, age, status as an individual with a disability, protected veteran status, genetic information, or other protected classes under the law. For additional information please see the University's Notice of Nondiscrimination.

Job seekers in need of a reasonable accommodation to complete the application process should call 773-702-1032 or email equalopportunity@uchicago.edu with their request.

Where Science Gets Social.

AAAS.ORG/COMMUNITY

Member COMMUNITY
AAAS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

Wake Forest™ School of Medicine

The Department of Neurobiology and Anatomy at Wake Forest Medical School announces open positions for

Postdoctoral Training in Multisensory Processes

We seek strong candidates for postdoctoral training funded by an NIH T32 Training Grant. The training program provides a rich collaborative research environment that fosters interdisciplinary approaches to understanding how the brain integrates information from multiple senses to produce perception and adaptive behavior. Candidates with direct experience as well as those in related fields are encouraged to apply. Trainees will have access to any of 9 laboratories using human subjects and/or a variety of animal models rodents-primates with approaches spanning molecular/cellular to perceptual/behavioral. Fellowships are awarded on a competitive basis.

Applications including a current curriculum vitae or nominations should be sent to the Training Grant Director: Dr. Barry E. Stein bestein@wakehealth.edu, or to its Co-Directors: Dr. Terrence Stanford (stanford@wakehealth.edu and Dr. Dwayne Godwin dgodwin@wakehealth.edu. A description of the faculty and the program can be accessed via the website: http://graduate.wfu.edu/admissions/t32/training_tpmp.html. Please also submit your application and CV to either job opening **38704** or **38705** at **www.wakehealth.edu/jobssearch/**.

EOE/AA: Minorities/Females/Disabled/Vets.

ScienceCareers online @sciencecareers.org

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

Confused about your next career move?

Download Free Career Advice Booklets!

ScienceCareers.org/booklets

Who's the Top Employer for 2019?

Science Careers' annual survey reveals the top companies in biotech & pharma voted on by *Science* readers.

Read the article and employer profiles and listen to podcasts at sciencecareers.org/topemployers

Science 2019
TOP EMPLOYERS

By **Desiree Dickerson**

Strange dreams

Four years into my Ph.D. program, my hands started to go numb each night as I slept. I didn't think much of it at the time; it was just uncomfortable and weird. Then one night, while sound asleep, I somehow mistook my numb arm for a snake. I threw myself out of bed and woke in a panic—panting, trembling, heart racing—on my bedroom floor, poised to defend myself against this imaginary snake. A few nights later, my hand was a spider. Once again, I launched myself out of bed in utter panic. I later discovered that these nighttime events weren't simply strange dreams; they were a direct result of grad school stress.

At the time, I was working toward my Ph.D. and training to be a clinical psychologist. I spent half my time sitting in a dark room listening to the "pop pop" of neurons firing as rats explored mazes and the other half seeing patients, helping everyday people work through anxiety, depression, stress, and burnout.

Juggling the Ph.D. and the clinical program caused the perfectionist in me to run riot. There was never enough time to immerse myself fully in either world. I oscillated between striving to be the best and, when I couldn't, struggling to find motivation to work at all. The clinical exam loomed large in my mind and played into my worst fear: being exposed in front of a panel of experts as the fraud I believed I was. I lived with a constant sense of impending doom.

"I lived with a constant sense of impending doom."

During lunch, I'd sit around the table in the department staff room with Ph.D. student and postdoc friends, joking about my nighttime attacks. "You thought your hand was a snake?" they'd say. Then we'd laugh, swapping stories about the sorry state we were in. At no point did I think I should see a doctor. It simply made for a funny story.

Later that year, while discussing a patient's symptoms with a neurologist, I realized that my anxiety could be the direct cause of the numbness in my hands. The perpetual stress was triggering tension in my shoulders, and that, in turn, was putting pressure on my nerves and cutting off the circulation to my extremities.

There was a certain irony in that realization. I'd spent years learning about psychological tools for dealing with stress and anxiety. But until then it never occurred to me that I, too, was experiencing an abnormal level of stress and anxiety—or that I might benefit from the tools I was using to help others. I had never seen my problems as anything other than reasonable responses to the pressures of academic life. When virtually all your peers are exhausted, stressed, and working overtime, how can you see your own experience as a cause for concern?

There were some exceptions—peers who had better work-life balance and didn't appear to be riddled with anxiety—but my friends and I persuaded ourselves that their more balanced approach was a failing, a lack of wanting. They clearly weren't on the same career path.

When I figured out why my hands were going numb, I turned to running to relieve my stress. That helped somewhat and my strange dreams went away, but it was a Band-Aid solution to what should have struck me as a more serious problem. A psychologist could have helped me manage my stressors more effectively and push back on the mounting workload.

I can't go back and change my approach to grad school. But I'm now in a position to show others an alternative path. I specialize in helping academics navigate similar challenges and take a healthier approach to dealing with the pressures of academia. I've worked with academics at all career stages, and I've seen firsthand how many of them harbor a chorus of inner voices demanding perfection and telling them that the sacrifices are necessary for success.

My message to those I work with is that the stereotype of the overworked, stressed-out academic is unhealthy and outdated. Productivity and well-being are not mutually exclusive. In fact, prioritizing your well-being can improve not only your productivity, but also your motivation, insight, creativity, and enjoyment. So, when your internal alarm bells send you a warning signal—or when snakes attack in the night—don't laugh; take action. ■

Desiree Dickerson is a neuroscientist and clinical psychologist based in Valencia, Spain.

ILLUSTRATION: ROBERT NEUBECKER

Published by AAAS