Lessons from the largest-ever trove of retracted papers *p. 390*

Identity and the pleasure of a partisan victory *p. 407*

Radiating pulses of light *p. 439*

Science

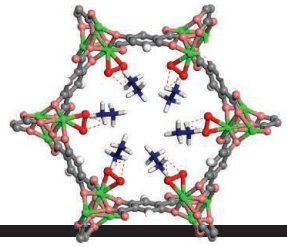
$15
26 OCTOBER 2018
sciencemag.org

AAAS

CHROMATIN LANDSCAPE

Illuminating the cancer epigenome *pp. 401 & 420*

CONTENTS



443
Holding ethane back

26 OCTOBER 2018 • VOLUME 362 • ISSUE 6413

NEWS

IN BRIEF

IN DEPTH

INSIGHTS

PERSPECTIVES

POLICY FORUM

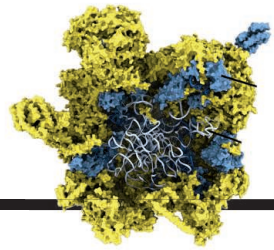
BOOKS *ET AL.*

LETTERS

FEATURE

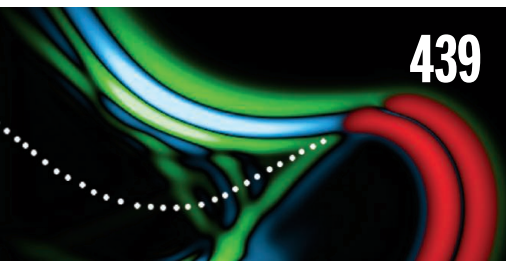
CREDITS (FROM TOP): *LI ET AL.*; DAVIDE BONAZZI/SAL ZMANART

CONTENTS



422 A complex ribosome

26 OCTOBER 2018 • VOLUME 362 • ISSUE 6413



439

RESEARCH

IN BRIEF

REVIEW

RESEARCH ARTICLES

REPORTS

DEPARTMENTS

ON THE COVER

Artist's rendering of chromatin (coiled structures) and DNA regulatory elements (yellow clusters) inside a cell's nucleus. The Cancer Genome Atlas (TCGA) Analysis Network profiled the complex accessible chromatin landscape in 410 cancer samples from 23 cancer types, substantially expanding the catalog of known DNA regulatory elements in cancer genomes and shedding light on how genetic mutations and epigenetic regulators work together to promote tumor development. See pages 401 and 420. *Image: V. Altounian*/Science

CREDITS (FROM TOP): HENSTRIDGE *ET AL.*

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. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $1808; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery): $89. 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; 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 is granted by AAAS to libraries and others who use Copyright Clearance Center (CCC) Pay-Per-Use services provided that $35.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

Published by AAAS

Science
AAAS
1200 New York Avenue, NW, Washington, DC 20005
Clarendon House, Clarendon Road, Cambridge, UK CB2 8FH

SUBSCRIPTION SERVICES For change of address, missing issues, new orders and renewals, and payment questions: 866-434-AAAS (2227) or 202-326-6417, FAX 202-842-1065. Mailing addresses: AAAS, P.O. Box 96178, Washington, DC 20090-6178 or AAAS Member Services, 1200 New York Avenue, NW, Washington, DC 20005

INSTITUTIONAL SITE LICENSES 202-326-6730 **REPRINTS:** Author Inquiries 800-635-7181 **COMMERCIAL INQUIRIES** 803-359-4578 **PERMISSIONS** 202-326-6765, permissions@aaas.org **AAAS Member Central Support** 866-434-2227 www.aaas.org/membercentral.

INFORMATION FOR AUTHORS See www.sciencemag.org/authors/science-information-authors

EDITORIAL

Imagine a world without facts

Editor-in-Chief, Science *Journals.*
jberg@aaas.org

When I was a young faculty member, my laboratory was deeply engaged in studies of zinc-containing proteins. One evening, after our two sons had been tucked into bed, I was watching The Simpsons and the episode began, uncharacteristically in black and white, with a man trying unsuccessfully to start his car. Another man says, "You said that you wanted to live in a world without zinc…" and explains the essential roles that zinc plays in various machines and expands on the frightening and dysfunctional scenario. I was genuinely concerned that I was hallucinating and wondered whom I could call to confirm what I thought I had seen. We are now living in a world where the reality of facts and the importance of scientific inquiry and responsible journalism are questioned with distressing frequency. This trend needs to be called out and arrested; the consequences of allowing it to continue are potentially quite damaging.

"THIS CREATES A WORLD OF IGNORANCE… CAUSING SUBSEQUENT DISCUSSIONS TO PROCEED WITHOUT MUCH FOUNDATION…"

Facts are statements that have a very high probability of being verified whenever appropriate additional observations are made. Thus, facts can be reliably used as key components in interpreting other observations, in making predictions, and in building more complicated arguments. For example, consider the proposition that the law of universal gravitation (LUG) accurately predicts the behavior of matter. The LUG is considered a fact given the vast and different sorts of observations that support it. The power of such accepted facts can be illustrated by the concept that the universe contains a large amount of "dark matter." Astronomer Vera Rubin obtained evidence that the detailed rotation of galaxies was not consistent with the LUG and other laws of physics, leading to two interpretations: Either the LUG was not correct in this context, or additional, large amounts of mass were present but not observable. Because of the wide support and acceptance of the LUG, the dark matter hypothesis was taken seriously, and additional evidence was interpreted in this context.

Consider the present "post-fact" world in this context. The lack of acceptance and cynical or ignorant questioning of well-documented evidence erode the perception that many propositions are well-supported facts, weakening the foundation on which many discussions and policies rest. Under these circumstances, numerous alternatives appear to be equally plausible because the evidence supporting some of these alternatives has been discounted. This creates a world of ignorance where many possibilities seem equally likely, causing subsequent discussions to proceed without much foundation and with outcomes determined by considerations other than facts.

Overly discounting information from appropriately trained researchers based on well-conducted studies, or from well-qualified journalists who pursue information with good practices that include interactions with multiple independent sources, can falsely restrict the available evidence. If evidence is judged without regard for its methodological basis, as well as a thorough assessment of sources of bias, unreliable conclusions may be drawn. If shoddy evidence is accepted, false interpretations may appear to be plausible even though they lack substantial evidentiary support. If robust evidence is undervalued or ignored, excess uncertainty will remain even when some propositions should be considered well established.

To avoid sliding further into a world without facts, we must articulate and defend the processes of evidence generation, evaluation, and integration. This includes not only clear statements of conclusions, but also clear understanding of the underlying evidence with recognition that some propositions have been well established, whereas others are associated with substantial remaining uncertainty. We should acknowledge and accept responsibility for, but not exaggerate, challenges within the scientific enterprise. At the same time, we should continue to call out statements put forward that are factually incorrect with reference to the most pertinent evidence. Without taking these steps forcefully, we risk living in a world where many things do not work as well as we need them to.

–Jeremy Berg

PHOTO: TERRY CLARK

10.1126/science.aav7494

> “There can be no genetics-based support for claiming one group as superior to another.”
>
> **The American Society of Human Genetics**, in a statement last week on what it calls a “societal resurgence” of using genetic concepts to bolster claims of white supremacy.

IN BRIEF

Edited by **Catherine Matacic**

BOUND FOR MERCURY

BepiColombo blasts off

BepiColombo, the European-Japanese mission to Mercury, began its 7-year journey to the solar system’s innermost planet with a 20 October launch. Riding atop an Ariane 5 rocket, the spacecraft blasted into the night sky above Europe’s spaceport at Kourou in French Guiana (above); half an hour later, it signaled it was safe in Earth orbit. In December, after system and instrument tests, controllers will turn on the spacecraft’s ion thrusters and begin a series of planetary encounters that will use gravity to boost it on its way. The first, with Earth, will be in April 2020, followed by two with Venus and six with Mercury itself. Finally, in December 2025, BepiColombo will enter into orbit around Mercury, and its two spacecraft will separate: Japan’s probe will study Mercury’s magnetosphere, and Europe’s orbiter will focus on the planet’s surface and interior, hoping to unravel the mysteries unsolved by previous probes, including why Mercury has an oversize iron core and volatile elements on its surface, and why its northern and southern hemispheres are so different (*Science*, 5 October, p. 11).

Cherished skull recovered

ARCHAEOLOGY | More than 40 years after it was unearthed from a cave in southeastern Brazil, the skull of Luzia, believed to be among the oldest human remains in the Americas, has been “rediscovered” in the scorched rubble of the country’s National Museum in Rio de Janeiro. The skull was stored in a metal case, within a metal cabinet, which protected it during the fire that destroyed the museum on 2 September. When researchers recovered the skull last week, it was broken and scarred, but in good enough shape to be reconstructed. The 11,500-year-old skull, which offers a glimpse of an early South American, was a celebrity in the museum’s 20-million-item collection, most of which was destroyed in the fire. “It was like a member of the family coming back to us,” said anthropologist Claudia Carvalho, who is overseeing recovery efforts at the museum.

Ebola efforts slowed

INFECTIOUS DISEASE | A deadly militia attack in Beni, Democratic Republic of the Congo, has once again set back efforts to contain an Ebola outbreak in the region, which had sickened 244 people and killed at least 122 by press time. On 20 October, according to government forces, rebels killed 15 civilians and kidnapped a dozen children in the city. Continuing unrest forced response team workers in the area to suspend their disease surveillance and vaccination efforts for several days. The World Health Organization decided on 17 October that the outbreak, though serious, does not yet constitute a Public Health Emergency of International Concern.

Embattled physicist to retire

#METOO | Lawrence Krauss, 64, the celebrity physicist who was demoted in July after Arizona State University (ASU) investigators concluded he grabbed a woman’s breast at a 2016 conference, announced on 21 October that he will retire in May 2019. His departure is part of a 19 October legal settlement with Tempe-based ASU, under which the university will drop procedures that could have led to his termination.

PHOTO: ESA/S. CORVAJA

Krauss maintains his innocence, writing on Facebook and Twitter, "I have never harassed or assaulted anyone." But ASU Provost Mark Searle determined in July that the breast-grabbing incident, seen by two onlookers, "created an unwelcome environment for academic pursuits" and violated the university's code of conduct. Under the terms of the settlement, some donors to the successful science popularizing program that Krauss founded, the Origins Project, can also request refunds.

Stem cell clinics are fined

BIOMEDICINE | The U.S. Federal Trade Commission (FTC) has taken its first enforcement action against a stem cell clinic for allegedly marketing unproven therapies. FTC targeted two California-based companies, Regenerative Medical Group, Inc. and Telehealth Medical Group, Inc., that advertised "amniotic stem cell therapy" as a remedy for autism, cerebral palsy, heart disease, and other serious illnesses. But FTC said there was no clinical evidence that the cells, derived from the amniotic fluid of women undergoing cesarean sections, were beneficial. In an 18 October settlement, the agency proposed a $3.31 million penalty that it said would be suspended if the firms' owner, Bryn Jarald Henderson, pays the agency $525,000, which may go toward refunds to customers.

India's CSIR gets new head

LEADERSHIP | For the first time since 1984, an outsider has taken the helm at India's Council of Scientific and Industrial Research (CSIR), a $1 billion public organization that operates a network of 40 research labs around the country. Shekhar Mande, a structural biologist who previously led the National Centre for Cell Science in Pune, India, took over as director-general on 16 October, with a mandate to make the organization self-financing—primarily through industry contracts. "CSIR is in urgent need of revitalization and only an outsider can bring in the new vigor needed to steer CSIR in a fast-changing India," says Ramakrishna Ramaswamy, a systems biologist and president of the Indian Academy of Sciences in Bengaluru.

New look at solar geoengineering

CLIMATE SCIENCE | Citing inaction by the world's countries to halt climate change, the National Academies of Sciences, Engineering, and Medicine (NASEM) is planning a new study of solar-radiation

PHOTO: J. ZHANG/INSTITUTE OF VERTEBRATE PALEONTOLOGY AND PALEOANTHROPOLOGY, BEIJING

PALEONTOLOGY

Is this the world's oldest bird lung?

Scientists have found what they think are the first preserved lungs from an ancient bird—traces of tissue in a 120-million-year-old fossil from China. The speckled white mass enmeshed in the ribs of *Archaeorhynchus spathula* (above) may offer new insights into how birds evolved the uber-efficient respiratory system that lets them soar. Like a modern bird's lung, the white mass is divided in two, flanked by what look like air sacs, and laced with tiny air capillaries that would have helped *A. spathula* absorb copious amounts of oxygen, says paleontologist Jingmai O'Connor of the Institute of Vertebrate Paleontology and Paleoanthropology in Beijing, who presented the work at the Society of Vertebrate Paleontology meeting last week in Albuquerque, New Mexico, and online in the *Proceedings of the National Academy of Sciences* on 22 October. Not everyone is convinced the white mass is a lung rather than a mineral artifact, but other researchers say O'Connor makes a good case.

management, a form of geoengineering that could temporarily cool the planet by reflecting sunlight. The project, announced 16 October, will produce a report on the environmental consequences of injecting particles into the upper atmosphere to reflect the sun's heat and offset continued global warming. It will also explore the international governance required to manage this technology. Unlike a similar 2015 report by NASEM, which was sponsored by U.S. federal agencies including NASA and the Department of Energy, the new study is sponsored by private foundations, though federal support is possible down the road.

WHO decries 'virginity testing'

PUBLIC HEALTH | The World Health Organization (WHO) and other United Nations agencies this month called for the immediate elimination of "virginity testing," used in more than 20 countries, including Afghanistan and South Africa, to determine women's rights to education, employment, marriage, and even freedom in nations where premarital sex is criminalized. The testing, which usually involves an inspection of the hymen to determine whether a woman has had vaginal intercourse, has no scientific or clinical basis, WHO said. The organization added that the test, also commonly used in criminal proceedings to determine whether a woman has been raped, violates human rights, and it calls on health providers to refuse to perform it.

Custom drug slows brain disease

GENE THERAPY | In less than a year, researchers have gone from sequencing a sick girl's genome to tailormaking a drug that appears to have halted her fatal disease. Seven-year-old Mila Makovec has Batten disease, a degenerative brain disorder. Timothy Yu and co-workers at Boston Children's Hospital found that Makovec had an extra piece of DNA in the noncoding part of a gene called *CLN7*, which normally creates a protein that helps cells clear waste products. The researchers designed a molecule called an oligonucleotide that masked the extra DNA and restored protein production in her culture cells; they then began to treat her with the drug in January. She is still blind and unable to speak or walk without help, but she now has fewer seizures and is clinically stable. Yu, who described the work last week at the annual meeting of the American Society of Human Genetics in San Diego, California, hopes the same approach can treat other genetic diseases caused by this type of mutation.

The last day of Pompeii in Italy may have been some 2 months later than archaeologists thought.

IMAGE: KARL BRULLOV/WIKIMEDIA COMMONS

ARCHAEOLOGY

Ancient inscription shifts date of Pompeii's destruction

New excavations of the Pompeii ruins in Italy have unearthed a charcoal inscription that calls into question the long-established date of Mount Vesuvius's eruption. Based on Pliny the Younger's eyewitness account, the cataclysmic event was supposed to have taken place on 24 August 79 C.E. But the inscription, likely made during a building renovation, includes the equivalent of the date 17 October 79 C.E. That suggests the explosion that buried the inscription in ash, preserving it, happened in late October. Hints of a later date had already come from previously discovered artifacts, including heavy garments and fall-harvested foods like pomegranates, walnuts, and grapes. The new date, bioarchaeologists say, could also shake up how preserved bodies from Pompeii are used to study the disease ecology of the time period, because many illnesses were seasonal.

Geography shapes birth canals

PHYSIOLOGY | People come in all different shapes, and so, too, do birth canals, according to a new study. An examination of 348 human pelvises from all over the world revealed certain geographic patterns: Women from sub-Saharan African and some Asian populations, for example, had the deepest birth canals, whereas Native American women had the widest ones. Europeans had birth canals whose upper sections were more markedly oblong than those of other populations. The authors suggest such data, published on 23 October in the *Proceedings of the Royal Society B*, could help inform obstetric practices around the world, as the differences in shape can lead to variations in childbirth.

Volunteers log 30,000 star images

SPACE SCIENCE | Citizen scientists have done for astronomy what many of us never find the time to do for ourselves: Prep old photos for digitizing. As part of a project called Astronomy Rewind, volunteers examined more than 30,000 celestial images from the journals of the American Astronomical Society, some more than a century old, for axes to orient them. What organizers had expected to take months, the enthusiastic amateurs did in a few days. The project's next phase, which began earlier this month, will use the axes logged by the volunteers to fill in each image's location, orientation, and scale. Images will then be uploaded to the WorldWide Telescope repository, so astronomers can study how the skies have changed since their predecessors wore frock coats.

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

Small and social, marmosets offer advantages for researchers.

PHOTO: TOM LANDERS/*THE BOSTON GLOBE*/GETTY IMAGES

ANIMAL RESEARCH

U.S. labs clamor for marmosets

Shortage develops as new transgenic models for neurological diseases stoke interest

By **Kelly Servick**, *in Washington, D.C.*

A hand-size monkey called *Callithrix jacchus*—the common marmoset—is in great demand in labs and yet almost nowhere to be found. Marmosets' small size, fast growth, and sophisticated social life were already enough to catch the eye of neuroscientists. They've now been genetically engineered to make their brains easier to image and to serve as models for neurological disorders such as autism and Parkinson's. The problem: "There are just no monkeys," says Cory Miller, a neuroscientist at the University of California, San Diego.

At a meeting here this week, convened by the National Academies of Sciences, Engineering, and Medicine's (NASEM's) Institute for Laboratory Animal Research, neuroscientist Jon Levine, who directs the Wisconsin National Primate Research Center at the University of Wisconsin in Madison, likened the surge in demand to "a 10-alarm fire that's about to be set." In response, the National Institutes of Health (NIH) plans to launch funding to expand marmoset research. And established marmoset researchers, including Miller, are working together to help new labs get animals.

When Miller's lab started to work with marmosets in 2009, many colleagues who studied macaques—the most popular genus of research monkey—didn't even know that marmosets were monkeys, he remembers. "They were like, 'Is it those chipmunks that were in the Rocky Mountains?'" (They were thinking of marmots.)

Now, he says, "All of those people want marmosets." In a survey, Miller and colleagues found that the number of U.S. marmoset research colonies jumped from eight in 2009 to 27 today, totaling 1900 marmosets across about 40 principal investigators.

Among monkeys, marmosets are known for cooperative social behavior: They call to each other in back-and-forth conversations, and mated pairs share responsibility for rearing young. They're smaller and easier to house than rhesus macaques, and they give birth twice a year versus once every year or two, aiding multigeneration genetic experiments. Because marmosets mature and age more quickly than bigger monkeys, they speed up studies of diseases that affect development and aging. And a marmoset's brain is less furrowed than a macaque's, which makes it easier to image or record activity from its surface.

Enthusiasm for marmosets surged in 2009, when they became the first primates shown to pass a genetic modification to offspring in their sperm and eggs. A team at the Central Institute for Experimental Animals (CIEA) in Kawasaki, Japan, injected embryos with the gene for a fluorescent protein. The skin and hair of the resulting animals shone green under ultraviolet light.

A series of transgenic marmosets followed—many from CIEA geneticist Erika Sasaki and neuroscientist Hideyuki Okano of Keio University in Tokyo. On 5 November at the Society for Neuroscience meeting in San Diego, their teams will present updates on two transgenic efforts: marmosets with genetic mutations that in humans are linked to Parkinson's disease and the neurodevelopmental disorder Rett syndrome. Researchers hope that by watching disease progress in a marmoset while analyzing its brain, they can lay bare mechanisms that cause illness in people—and maybe find and test new therapies.

Japanese research got a leg up in 2014 with a 40 billion yen ($350 million) government initiative to map the marmoset brain. But several U.S. labs now have transgenic primates under development. In 2016, a team at NIH's National Institute of Neurological Disorders and Stroke, with Sasaki, created marmosets with brain cells that fluoresce when excited—a potential tool for monitoring neural activity. And in April, the first marmoset with a mutation in the gene *SHANK3*—implicated in some cases of autism—was born at the Massachusetts Institute of Technology (MIT) in Cambridge.

Making transgenic monkeys requires a large colony, in part because females im-

planted with manipulated embryos don't always get pregnant. Guoping Feng, who leads the MIT project, estimates the ideal size is at least 300 animals, far more than a single U.S. facility can breed. (Feng's group has gradually built up a colony of about 200.) When the new transgenic models become widely available—likely in the next few years—labs hoping to use them may also need their own animals for breeding. Attendees at this week's meeting also discussed ways to maintain genetic diversity within the U.S. marmoset population.

But the supply of new marmosets is limited. An international agreement restricts the export of wild animals from their native Brazil. And importing animals from breeding facilities in Asia is "really, really difficult," Feng says. Most airlines, facing pressure from animal rights groups, have stopped carrying research animals (*Science*, 5 October, p. 15).

Already, public resistance to nonhuman primate research is prompting researchers to tread carefully. Increasing interest in marmoset research is "concerning to us," says Kathleen Conlee, vice president of animal research issues at the Humane Society of the United States here. It's especially problematic, she says, to genetically design animals that will become ill.

But scientists see no substitute for primates in some studies. "When it comes to [studying] cognitive processes and other complex behaviors, some things you just need to do in a primate model," Joshua Gordon, director of NIH's National Institute of Mental Health in Bethesda, Maryland, said at a 4 October NASEM meeting on genetically engineered nonhuman primates. The study of mental illness requires an understanding of brain structures that don't exist in rodents, he added. But such research must consider "the degree to which primate experiments are acceptable to the general public," he said.

Next year, Gordon's agency plans to announce funding opportunities to support centralized infrastructure for marmoset research. Although details are hazy, the funding might bring in new marmosets, expand or establish breeding colonies, or advance transgenic projects, he said. Its money could come from the federal Brain Research through Advancing Innovative Neurotechnologies Initiative or NIH's Blueprint for Neuroscience Research.

In the meantime, labs are improvising. Last month, several investigators launched a virtual pool, to which existing marmoset colonies will contribute 10% of their animals per year for new investigators to buy or inherit. It's a stopgap to keep up momentum in the field, Miller says, "because it's kind of a once-in-a-career opportunity." ■

PHOTO: VCG/GETTY IMAGES

SUPERCOMPUTERS

China fosters competition in its race to exascale computing

Three teams are developing prototypes using Chinese chips

By **Dennis Normile**, *in Tianjin, China*

In a cavernous room just off the marble floored lobby of China's National Supercomputer Center of Tianjin stand more than 100 wardrobe-size black and gray metal cabinets, arranged in ranks like a marching army. They contain the Tianhe-1A supercomputer, which 8 years ago became the first Chinese machine to reign, briefly, as the world's fastest computer, running at 2.57 petaflops (or quadrillion floating point operations per second). But just upstairs from Tianhe-1A—and off-limits to visitors—is a small prototype machine that, if successfully scaled up, could push China to the top of the rankings again. The goal is a supercomputer capable of 1 exaflop—1000 petaflops, five times faster than the current champion, the Summit supercomputer at Oak Ridge National Laboratory in Tennessee.

China is vying with the United States, Europe, and Japan to plant its flag in this rarefied realm, which will boost climate and weather modeling, human genetics studies, drug development, artificial intelligence, and other scientific uses. But its strategy is unique. Three teams are competing to build China's machine; the Tianjin prototype has rivals at the National Supercomputing Center in Jinan and at Dawning Information Industry Co., a supercomputer manufacturer in Beijing. The Ministry of Science and Technology (MOST) will probably select two for expansion to exascale by the end of the year. The approach is a chance to spur innovation, says Bob Sorensen, a high-performance computing analyst at Hyperion Research in St. Paul. It "encourages vendors to experiment with a wide range of designs to distinguish themselves from their competitors," he says.

China may not be first to reach this computing milestone. Japan's Post-K exascale computer could be running in 2020. The United States is aiming to deploy its first exascale system at Argonne National Laboratory in Lemont, Illinois, in 2021 (*Science*, 9 February, p. 617). The European Union is ramping up its own program. China is aiming for 2020, but the date may slip.

Being first is not China's only goal, however. Having three competing teams will ensure broad-based technological advancement in computer chips, operating software, networking, and data storage technologies, says Meng Xiangfei, a physicist leading exascale application R&D for the center here. Building domestic capacity is particularly important for central processing units (CPUs) and specialized chips called accelerators, which boost a computer's performance. China relied on U.S.-made Intel CPUs for several generations of supercomputers, says Jack Dongarra, a computer scientist at the University of Tennessee in Knoxville, but in 2015, the U.S. government barred the export of certain chips

Tianhe-1A was the world's fastest computer in 2010. Its successor is being developed in the same building.

for security reasons. That move "provoked the Chinese government to make a heavy investment" in processors, he says. All three exascale prototypes use chips made in China.

The three-team strategy also allowed MOST to share costs with regional governments, which hope that a leading edge supercomputer will spur technological development and lure institutes and businesses. Qian Depei, a computer scientist at Beihang University in Beijing who serves on MOST's exascale evaluation team, says the prototypes cost about $9 million each; MOST put up half and the rest came from local sources.

The prototypes have faced a battery of tests for speed, stability, and energy consumption plus trial runs of software from different application areas, but the results are "very secret," Meng says. The final budget is also unclear, though at the outset a governmental advisory committee estimated one exascale computer would cost 2 billion to 3 billion yuan ($288 million to $432 million).

Even after the two winners are announced, Qian says the third team will probably remain involved so the expertise they've acquired is not wasted. Scaling up the prototypes, which operate in the range of 3 petaflops, will mean interconnecting enough CPUs and accelerators to reach an exaflop, refining the liquid cooling systems needed to remove heat and improve efficiency, and perfecting the operating software needed for the massively parallel arrangement of processors to work together.

China once lagged in developing application software needed to do interesting science with supercomputers, but it has been catching up, Meng says. For the past 2 years, Chinese groups have won the Gordon Bell Prize presented annually by the Association for Computing Machinery for innovations in applying high-performance computing to science, engineering, and large-scale data analytics. Chinese scientists are now working on new applications, says Yang Meihong, director of the Jinan center. For example, going to exascale will allow a dramatic improvement in the spatial resolution of global atmospheric models, "which will be greatly significant for a deeper understanding of the mechanisms of climate change," she says.

The United States still dominates among the truly powerful supercomputers used for research, with 21 systems in the top 50 to China's two. But scientists play down the ranking's importance. "Having the top 500 No. 1 supercomputer—that's pretty good, but that's not the goal," Qian says. "The real measure should be what kind of new science we have as a result of these computers," Dongarra says. ■

HUMAN GENETICS

Giant study links DNA to same-sex experiences

People who reported having at least one same-sex partner are more likely to share certain genetic variants

By **Michael Price**, *in San Diego, California*

How genes influence sexual orientation has sparked debate for at least a quarter-century. But geneticists have had only a handful of underpowered studies to address a complex, fraught, and often stigmatized area of human behavior. Now, the largest-ever study of the genetics of sexual orientation has revealed four genetic variants strongly associated with what the researchers call nonheterosexual behavior. Some geneticists are hailing the findings as a cautious but significant step in understanding the role of genes in sexuality. Others question the wisdom of asking the question in the first place.

Andrea Ganna of the Broad Institute in Cambridge, Massachusetts, and Harvard Medical School in Boston, and colleagues examined data from hundreds of thousands of people who provided both DNA and behavioral information to two large genetic surveys, the UK Biobank study and the private genetics firm 23andMe. The team analyzed DNA markers from people who answered either "yes" or "no" to the question, "Have you ever had sex with someone of the same sex?" In total, they identified 450,939 people who said their sexual relationships had been exclusively heterosexual and 26,890 people who reported at least one same-sex experience.

> ***"We know almost nothing about the genetics of sexual behavior, so anywhere is a good place to start."***
> **Andrea Ganna**, Broad Institute

In Ganna's talk last week at the annual meeting of the American Society of Human Genetics here, he emphasized that the researchers were cautious about exploring sexual behavior that is still illegal in many countries, and that they tried to frame their research goals carefully "to avoid a fishing expedition." The team, which includes behavioral scientists, preregistered their research design and also met regularly with members of the lesbian, gay, bisexual, transgender, or questioning (LGBTQ) community to discuss and share results. Ganna acknowledged that what the team calls nonheterosexual behavior includes "a large spectrum of sexual experiences, that go from people who engage exclusively in same-sex behavior to those who might have experimented once or twice."

The researchers performed a genome-wide association study (GWAS) in which they looked for specific variations in DNA that were more common in people who reported at least one same-sex sexual experience. They identified four such variants on chromosomes seven, 11, 12, and 15.

Two variants were specific to men who reported same-sex sexual experiences. One, a cluster of DNA on chromosome 15, has previously been found to predict male-pattern baldness. Another variant on chromosome 11 sits in a region rich with olfactory receptors. Ganna noted that olfaction is thought to play a large role in sexual attraction.

A much smaller 1993 study, which used a different kind of association technique known as a genetic linkage study, had suggested that a stretch of DNA on the X chromosome was linked to inherited homosexuality. In the new GWAS, that stretch was not found to be associated with reported same-sex behavior. But the lead author of the earlier study, Dean Hamer, then of the National Institutes of Health in Bethesda, Maryland, praised the new work. "It's important that attention is finally being paid [to the genetics of sexual orientation] with big sample sizes and solid institutions and people," he said. "This is exactly the study we would have liked to have done in 1993."

The four newly identified genetic variants also were correlated with some mood and mental health disorders. Both men and women with the variants were more likely to have experienced major depressive disorder and schizophrenia, and women were more likely to have bipolar disorder. Ganna stressed that these findings should not be

taken to mean that the variants cause the disorders. Instead, it "might be because individuals who engaged in nonheterosexual behavior are more likely to be discriminated [against], and are more likely to develop depression," he said.

Ganna noted that the correlation with schizophrenia and risk-taking behavior was more pronounced in the UK Biobank participants, who tend to skew older than those in the 23andMe group. That could be because older generations faced more sexual discrimination than younger ones, Ganna said, noting that environment likely plays a significant role in which traits wind up correlating with sexual behavior.

Overall, he said the findings reinforce the idea that human sexual behavior is complex and can't be pinned on any simple constellation of DNA. "I'm pleased to announce there is no 'gay gene,'" Ganna said. "Rather, 'nonheterosexuality' is in part influenced by many tiny genetic effects." Ganna told *Science* that researchers have yet to tie the genetic variants to actual genes, and it's not even clear whether they sit within coding or noncoding stretches of DNA. Trying to pin down exactly what these DNA regions do will be among the team's difficult next steps.

"It's an intriguing signal," he said. "We know almost nothing about the genetics of sexual behavior, so anywhere is a good place to start." He added that the four genetic variants could not reliably predict someone's sexual orientation. "There's really no predictive power," he said.

Given the complexity of human sexual behavior, much of which is not captured in the study questions, biomedical informatics graduate student Nicole Ferraro from Stanford University in Palo Alto, California, questioned the work's utility. She and fellow biomedical sciences grad student Kameron Rodrigues said the study didn't do enough to explore the nuances of how a person's sexual identity may differ from their sexual behavior, and they worried that the research could be used to stigmatize members of the LGBTQ community. "It just seems like there's no benefit that can come from this kind of study, only harm," Rodrigues said.

The abstract for Ganna's talk referenced another provocative result: Heterosexual people who possess these same four genetic variants tend to have more sexual partners, suggesting associated genes might confer some mating advantage for heterosexuals. That could help explain why these variants might stick around in populations even if people attracted to the same sex tend to have fewer children than heterosexuals. Ganna did not touch on that finding in his talk, citing lack of time.

That was probably a wise choice, geneticist Chris Cotsapas at the Yale School of Medicine said, because the evolutionary implications haven't been firmed up. "People are going to oversimplify it to say, 'Gay genes help straight people have more sex,' and it's really not that simple," he said.

Overall, the findings were "very carefully, cautiously presented," Cotsapas said, and represent a good start for geneticists charting the complexities of human sexuality. ■

With reporting by Jocelyn Kaiser. Michael Price is a science journalist in San Diego, California.

GLOBAL WARMING

Youth climate trial showcases science

Government won't dispute climate change, but will cast doubt on claimed harms

By **Julia Rosen**

Next week, barring a last-minute intervention by the Supreme Court, climate change will go to trial for just the second time in U.S. history. In a federal courtroom in Eugene, Oregon, 21 young people are scheduled to face off against the U.S. government, which they accuse of endangering their future by promoting policies that have increased emissions of carbon dioxide (CO_2) and other planet warming gases. The plaintiffs aren't asking for monetary damages. Instead, they want District Judge Ann Aiken to take the unprecedented step of ordering federal agencies to dramatically reduce the amount of CO_2 in the atmosphere.

Government attorneys are not expected to challenge the scientific consensus that human activities, including the burning of fossil fuels, cause global warming. But the outcome could hinge, in part, on how Aiken weighs other technical issues. Each side has recruited a roster of high-profile scientists and economists, including Nobel laureates, to bolster their argument. "It's clearly going to be a battle of the experts," says Michael Gerrard, director of the Sabin Center for Climate Change Law at Columbia University, who is not involved in the case.

The civil trial will be a milestone in a hard-fought legal battle that began in 2015, when environmental groups joined with youth activists and retired NASA scientist James Hansen to push for climate action. The lawsuit rests on the novel argument that the government has knowingly violated the plaintiffs' rights to a "safe" climate by taking actions—such as subsidizing fossil fuels—that cause warming. Government lawyers under both President Donald Trump and former President Barack Obama have repeatedly tried and failed to have the case thrown out. Just last week, Aiken, who was

ILLUSTRATION: GARY WATERS/THEISPOT

Levi Draheim, a plaintiff in the lawsuit, at a 2017 rally in Washington, D.C.

PHOTO: ROBIN LOZNAK/OUR CHILDREN'S TRUST

appointed by former President Bill Clinton, denied a bid to stop the trial, which was set to open on 29 October. But the Supreme Court then froze the case while it considers the government's argument that the plaintiffs' lawsuit has fatal legal flaws.

If the trial proceeds, the youths' lawyers will have to persuade the judge that the government's actions have helped cause climate change; that the warming exacerbated storms, droughts, and wildfires; and that individual plaintiffs have suffered injuries as a result. One expert for the plaintiffs, climate scientist Kevin Trenberth of the National Center for Atmospheric Research in Boulder, Colorado, filed testimony based on his research that invokes principles of thermodynamics to explain how warming can amplify disasters. Hotter, dryer weather increases the risk of fires, he notes. And warmer air can hold more moisture, boosting rainfall by up to 20%—a factor he says worsened floods that affected plaintiffs living in Louisiana, Florida, and Colorado. "Once thresholds are crossed, things break, burn, or die!" he writes.

Defense expert John Weyant, a management scientist at Stanford University in Palo Alto, California, disagrees that weather extremes can be pinned on climate change. Existing science is not precise enough to attribute individual events and injuries to climate, he writes in his testimony, and Trenberth's method ignores confounding factors. For example, he argues, poor forest management and human-sparked blazes also cause severe fires.

These confounding factors are "true but irrelevant" to whether warming contributed to a particular weather extreme, says climate scientist Drew Shindell of Duke University in Durham, North Carolina. But Trenberth's approach is controversial, notes climate scientist Friederike Otto of the University of Oxford in the United Kingdom. (Neither scientist is involved in the case.) Researchers can now at least partly attribute extreme events to warming, Otto says, but Trenberth's method isn't able to rule out whether a particular event would have occurred even without climate change.

The two sides may also spar over health impacts. Some plaintiffs claim smoke from worsening wildfires has triggered their asthma attacks. Others say the ecological harm driven by warming has caused emotional distress. But doctors testifying for the defense say the plaintiffs can't point to medical evidence for these connections. Samantha Ahdoot, a pediatrician at Virginia Commonwealth University's Inova campus in Richmond who is not involved in the case, notes it's often impossible to determine the cause of specific ailments. But she says large-scale studies support claims that climate impacts can harm children's health.

Even if the youth prevail on scientific questions, Gerrard says, they'll have to persuade the judge that she can provide an effective remedy. In other recent (and so far unsuccessful) climate lawsuits, cities and allied plaintiffs have sought money from energy companies and others to help them address alleged harms. But the youth want Aiken to order the Environmental Protection Agency (EPA) and other agencies to revamp regulations, with the goal of reducing atmospheric concentrations of CO_2 to 350 parts per million (ppm), from about 410 ppm today.

The two sides disagree over whether the United States can reduce emissions as fast as the plaintiffs would like. But the government may also argue that any cuts would have a limited impact, because of the global nature of climate change. Other countries now produce roughly 88% of the world's greenhouse gas emissions, David Victor, a climate policy expert at the University of California, San Diego, notes in his testimony for the defense. Therefore, the solution requires international cooperation, he writes.

Courts have rejected that argument as a rationale for inaction in previous cases, Gerrard says. In the only other climate lawsuit to get to the trial stage—a landmark 2007 case that challenged EPA's refusal to regulate CO_2 from vehicles—the agency claimed doing so wouldn't matter because U.S. cars contributed just 6% of global emissions. But the Supreme Court disagreed.

Perhaps the biggest question at stake, observers say, is whether judges can—and should—set climate policy. The plaintiffs think so, calling the courts their "last resort," and Aiken appears open to the idea. "Federal courts too often have been cautious and overly deferential in the arena of environmental law, and the world has suffered for it," she wrote in a 2016 decision. But that could make the judge "essentially the carbon regulator of the entire United States economy," says attorney Justin Torres of King & Spalding in Washington, D.C., who handled the case as a government lawyer under Obama. That would be politically and practically unwise, he says, and the government has argued climate policy is best tackled by Congress and the White House.

Such questions will likely resurface during appeals no matter how this case turns out, Gerrard says. But even if the trial is halted, it has already forced the Trump administration to give ground on climate change. The case has required "the administration to make numerous admissions," he says, "and be much more specific in its scientific claims than it has had to be before." ■

Julia Rosen is a journalist based in Portland, Oregon.

In large enough numbers, caribou and other large grazers could slow the thawing of permafrost.

PHOTO: SYLVAIN CORDIER/GAMMA-RAPHO/GETTY IMAGES

ECOLOGY

Restoring lost grazers could help blunt climate change

Rewilded landscapes could be cooler, less fire prone

By **Elizabeth Pennisi**

Restoring reindeer, rhinoceroses, and other large mammals could help protect grasslands, forests, and tundra from catastrophic wildfires and other threats associated with global warming, new studies suggest. The findings give advocates of so-called trophic rewilding—reintroducing lost species to reestablish healthy food webs—a new rationale for bringing back the big grazers.

Rewilding offers "solutions to some of the important problems arising from climate change," says ecologist Jens-Christian Svenning of Aarhus University in Denmark, an editor of a special issue on the topic published this week in the *Philosophical Transactions of the Royal Society B*. "The scope for … beneficial spin-offs for human society is tremendous," adds co-editor Elisabeth Bakker, an ecologist at the Netherlands Institute of Ecology in Wageningen.

Rewilding is often associated with an ambitious proposal to restore large mammals, including even ice age mammoths, to a huge park in Russia (*Science*, 4 December 2015, p. 1148). But mammoth resurrection is still just a dream, and most rewilders are focused on restoring animals including giant tortoises, dam-building beavers, or herds of grazers.

Now, it seems rewilding could offer a climate bonus. As the planet has warmed, fire seasons have become 25% longer than they were 30 years ago, and more areas are experiencing severe blazes, notes ecologist Christopher Johnson of the University of Tasmania in Hobart, Australia. He and his colleagues combed the literature back to 1945 for data on habitats around the world that have lost or gained large grazers, to see whether they experienced change in fire frequency or intensity. They found studies of 14 ancient landscapes up to 43,000 years old that used fungal spores associated with dung as a proxy for herbivore abundance, charcoal as a proxy for fire frequency, and pollen to reveal past vegetation. In about half of the landscapes, fires increased and the vegetation changed after the herbivores disappeared, they report. The other studies saw no clear effect.

The researchers also examined records from three modern landscapes, including the 100-year-old Hluhluwe Imfolozi Park in South Africa. There, data show larger and more frequent fires occurred after managers culled or moved large grazers, including white rhinos, wildebeests, zebras, buffalos, and impalas. (Rangers moved the animals as part of tsetse fly eradication efforts or to reduce overgrazing.) In the case of white rhinos, fires averaged just 10 hectares when the animals were present—because they kept plants closely cropped, and their paths created fire breaks—but increased to an average of 500 hectares after the rhinos vanished.

Similarly, in Australian grasslands, Johnson's team found that herds of feral swamp buffalos have helped control wildfires by similar means. When the researchers examined the herbivore and fire history of the southwestern United States, other grazers—including pronghorn antelopes, desert bighorn sheep, bison, and even domestic cattle—seemed to have helped starve fires by eating the grasses that serve as fuel. Modeling studies, meanwhile, suggest some grazers reduce fire risks by disturbing the soil, which buries leaf litter and other flammable material. The bottom line, Johnson and his colleagues write, is that "vertebrates can have strong effects on fire regimes," and restoring large grazers could be a fairly effective control measure at a time when risks are growing.

Other studies suggest grazers can also help maintain tundra—the semifrozen, treeless ecosystem found in the Arctic and on high mountains. In the Arctic, rapidly warming temperatures are enabling trees and shrubs to invade the tundra. The woody plants amplify Arctic warming by absorbing heat and trapping a layer of snow that insulates the ground, keeping it warmer. The result is that the soil thaws and releases even more stored carbon and other warming gases.

Reintroducing large numbers of herbivores to browse on the shrubs could help stop this vicious cycle, write ecologists Johan Olofsson at Umeå University in Sweden and Eric Post at the University of California, Davis. The more species—muskoxen, moose, bison, or caribou (also known as reindeer)—the better. But they say the biggest benefit might come from rethinking existing hunting and development rules to create the densest and most diverse herds possible. Rewilding might be "one of the few ways humans in the Arctic can mitigate global warming, or at least its consequences," Olofsson says.

Others are skeptical. Like any ecosystem re-engineering effort, the long-term effects of rewilding are hard to anticipate, warns Joseph Bump, an ecologist at the University of Minnesota in St. Paul. Some modeling, for example, suggests increased Arctic grazing will lead to greater carbon release, not less. And creating Arctic herds big enough to make a difference could be difficult, says ecologist Andrew Tanentzap of the University of Cambridge in the United Kingdom. They could become "a drop in the bucket in the sea of melting permafrost."

Even the strongest rewilding advocates concede its limits. "It would be overly optimistic to claim rewilding as 'the solution' to climate change," Svenning says. "But [it] clearly can play a role." ■

CHEMISTRY

Molecular CT scan could speed drug discovery

Structure mapping works with vanishingly small samples in just minutes

By **Robert F. Service**

In chemistry, structure rules: A molecule's 3D shape determines how it behaves. But the two standard ways to map the structure of small organic molecules, such as pharmaceuticals, hormones, and vitamins, have drawbacks. Last week, two research teams reported they've adapted a third technique, a variant of electron microscopy commonly used to chart much larger proteins, to determine the precise shape of small organic molecules. The new technique works with vanishingly small samples, is blazing fast, and is surprisingly easy.

"I am blown away by this," says Carolyn Bertozzi, a chemist at Stanford University in Palo Alto, California, who wasn't involved in the work. "The fact that you can get these structures from [a sample] a million times smaller than a speck of dust, that's beautiful. It's a new day for chemistry."

The gold standard for determining chemical structures is x-ray crystallography. A beam of x-rays is fired at a pure crystal containing millions of copies of a molecule lined up in a single orientation. By tracking how the x-rays bounce, researchers can work out the position of every atom in the molecule. Crystallography can pinpoint atomic positions down to less than 0.1 nanometers, about the size of a sulfur atom. But the technique works best with fairly large crystals, which can be hard to make. "Just getting a crystal … can take weeks to months to years," says Brian Stoltz, an organic chemist at the California Institute of Technology in Pasadena.

The second approach, nuclear magnetic resonance (NMR) spectroscopy, doesn't require crystals. It infers structures by perturbing the magnetic behavior of atoms in molecules and then tracking their behavior, which changes depending on their atomic neighbors. But NMR requires a fair amount of material. And it's indirect, which can lead to mapping mistakes.

The new approach builds on a technique called electron diffraction, which sends an electron beam through a crystal and, as in x-ray crystallography, determines structure from diffraction patterns. It has been particularly useful in solving the structure of a class of proteins lodged in cell membranes. In this case, researchers first form tiny, 2D, sheetlike crystals of multiple copies of a protein wedged in a membrane.

But in many cases, efforts to grow the protein crystals go awry. Instead of getting single-membrane sheets, researchers end up with numerous sheets stacked atop one another, which can't be analyzed by conventional electron diffraction. And the 3D crystals can be too small for x-ray diffraction. So, Tamir Gonen, an electron crystallography expert at the University of California, Los Angeles (UCLA), and his team varied the technique: Rather than firing their electron beam at a static crystal, they continually rotated the crystal and tracked how the diffraction pattern changed. Instead of a single image, they got something more like a molecular computerized tomography scan. That gave them enough information to solve structures from crystals just 100 nanometers on a side, one-billionth the size of those needed for x-ray crystallography.

Gonen says because his interest was in proteins, he never thought much about trying his technique on anything else. But earlier this year, he moved to UCLA, where his new colleagues, along with Stoltz, wanted to see whether the approach would work with smaller organic molecules. The short answer is it did. On the preprint server ChemRxiv, the California team reported last week that with numerous samples, the method worked nearly every time, delivering a resolution on par with x-ray crystallography. The team could even get structures from mixtures of compounds and from materials that had never formally been crystallized. These results all came after just a few minutes of sample preparation and data collection.

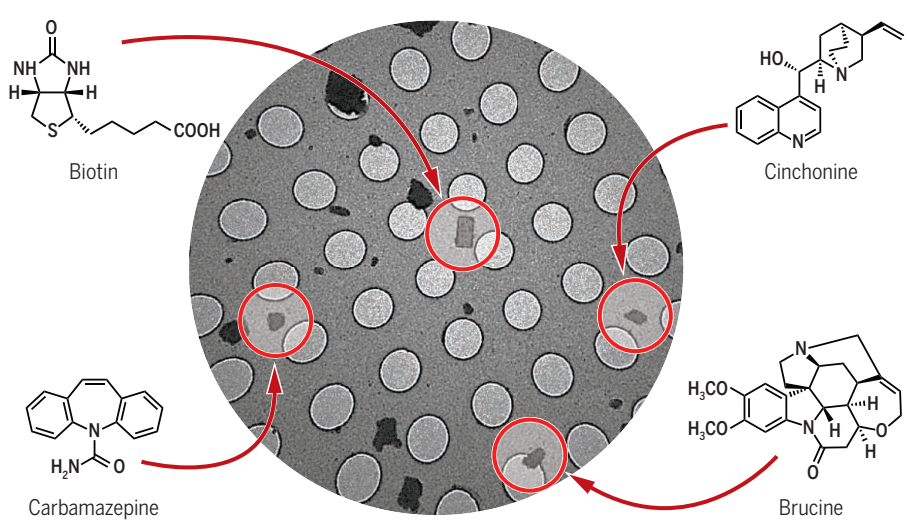
What's more, a collaboration of German and Swiss groups independently published similar results last week in *Angewandte Chemie International Edition* using essentially the same technique. That paper's first author, Tim Grüne, an electron diffraction expert at the Paul Scherrer Institute in Villigen, Switzerland, says his team found inspiration in work by chemists who use electron diffraction to study gems and other inorganic materials, not proteins. But that led them to much the same place.

"I've had dreams in my life where I'm looking through a microscope and I see a molecular model with balls and sticks," Bertozzi says. "They basically find some microcrystalline schmutz on an EM [sample holder], take some data, and there are the balls and sticks I dreamed about."

The technique could revolutionize fields inside and outside of research, Bertozzi and others say. Grüne notes that pharmaceutical companies build massive collections of potential new drugs. But few compounds form crystals big enough for x-ray crystallography. "This will remove a bottleneck and lead to an explosion of structures," Grüne says. ■

Structures from a mix of microcrystals

A new technique identified structures of four compounds from tiny crystals on an electron microscope slide.



GRAPHIC: C. JONES *ET AL.*, CHEMRXIV. (2018) ADAPTED BY A. CUADRA/*SCIENCE*

FEATURES

RETHINKING RETRACTIONS

The largest-ever database of retracted articles suggests the burgeoning numbers reflect better oversight, not a crisis in science

Published by AAAS

By **Jeffrey Brainard**; *Data analysis and graphics by* **Jia You**; *Illustrations by* **Davide Bonazzi**

Nearly a decade ago, headlines highlighted a disturbing trend in science: The number of articles retracted by journals had increased 10-fold during the previous 10 years. Fraud accounted for some 60% of those retractions; one offender, anesthesiologist Joachim Boldt, had racked up almost 90 retractions after investigators concluded he had fabricated data and committed other ethical violations. Boldt may have even harmed patients by encouraging the adoption of an unproven surgical treatment. Science, it seemed, faced a mushrooming crisis.

The alarming news came with some caveats. Although statistics were sketchy, retractions appeared to be relatively rare, involving only about two of every 10,000 papers. Sometimes the reason for the withdrawal was honest error, not deliberate fraud. And whether suspect papers were becoming more common—or journals were just getting better at recognizing and reporting them—wasn't clear.

Still, the surge in retractions led many observers to call on publishers, editors, and other gatekeepers to make greater efforts to stamp out bad science. The attention also helped catalyze an effort by two longtime health journalists—Ivan Oransky and Adam Marcus, who founded the blog Retraction Watch, based in New York City—to get more insight into just how many scientific papers were being withdrawn, and why. They began to assemble a list of retractions.

That list, formally released to the public this week as a searchable database, is now the largest and most comprehensive of its kind. It includes more than 18,000 retracted papers and conference abstracts dating back to the 1970s (and even one paper from 1756 involving Benjamin Franklin). It is not a perfect window into the world of retractions. Not all publishers, for instance, publicize or clearly label papers they have retracted, or explain why they did so. And determining which author is responsible for a paper's fatal flaws can be difficult.

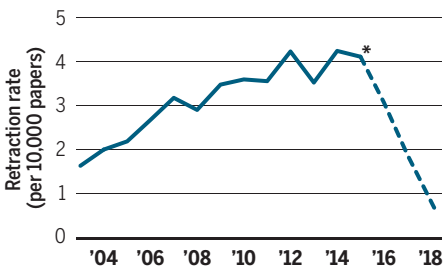
Still, the data trove has enabled *Science*, working with Retraction Watch, to gain unusual insight into one of scientific publishing's most consequential but shrouded practices. Our analysis of about 10,500 retracted journal articles shows the number of retractions has continued to grow, but it also challenges some worrying perceptions that continue today. The rise of retractions seems to reflect not so much an epidemic of fraud as a community trying to police itself.

Among the most notable findings:

Although the absolute number of annual retractions has grown, the rate of increase has slowed. The data confirm that the absolute number of retractions has risen over the past few decades, from fewer than 100 annually before 2000 to nearly 1000 in 2014. But retractions remain relatively rare: Only about four of every 10,000 papers are now retracted. And although the rate roughly doubled from 2003 to 2009, it has remained level since 2012. In part, that trend reflects a rising denominator: The total number of scientific papers published annually more than doubled from 2003 to 2016.

Retraction rate levels off

Although the number of retractions ballooned after 1997, the percentage of all papers retracted rose more slowly and leveled off after 2012.



*The rate appears to decline after 2015, but numbers are almost certainly incomplete because of delays in publication of retractions.

Much of the rise appears to reflect improved oversight at a growing number of journals. Overall, the number of journals that report retractions has grown. In 1997, just 44 journals reported retracting a paper. By 2016, that number had grown more than 10-fold, to 488. But among journals that have published at least one retraction annually, the average number of retractions per journal has remained largely flat since 1997. Given the simultaneous rise in retractions, that pattern suggests journals are collectively doing more to police papers, says Daniele Fanelli, a lecturer in research methods at the London School of Economics and Political Science who has co-written several studies of retractions. (The number per journal would have increased, he argues, if the growing number of retractions resulted primarily because an increased proportion of papers are flawed.)

"Retractions have increased because editorial practices are improving and journals are trying to encourage editors to take retractions seriously," says Nicholas Steneck, a research ethics expert at the University of Michigan in Ann Arbor. Scientists have kept the pressure on journals by pointing out flaws in papers on public websites such as PubPeer.

In general, journals with high impact factors—a measure of how often papers are cited—have taken the lead in policing their papers after publication. In 2004, just one-fourth of a sampling of high-impact biomedical journals reported having policies on publishing retractions, according to the *Journal of the Medical Library Association* (*JMLA*). Then, in 2009, the Committee on Publication Ethics (COPE), a nonprofit group in Eastleigh, U.K., that now advises more than 12,000 journal editors and publishers, released a model policy for how journals should handle retractions. By 2015, two-thirds of 147 high-impact journals, most of them biomedical titles, had adopted such policies, *JMLA* reported. Proponents of such policies say they can help journal editors handle reports of flawed papers more consistently and effectively—if the policies are followed.

Journals with lower impact factors also appear to be stepping up their standards,

About the data

With more than 18,000 entries overall, the database (retractiondatabase.org) created by the staff of the Retraction Watch blog contains more retractions from more journals than databases maintained by other organizations. The National Library of Medicine's PubMed database, for example, contains only biomedical journals. In addition, Retraction Watch staff members actively search for retraction notices instead of relying only on notices from journal publishers. Retraction Watch also uses a taxonomy it developed to record the reason for each retraction.

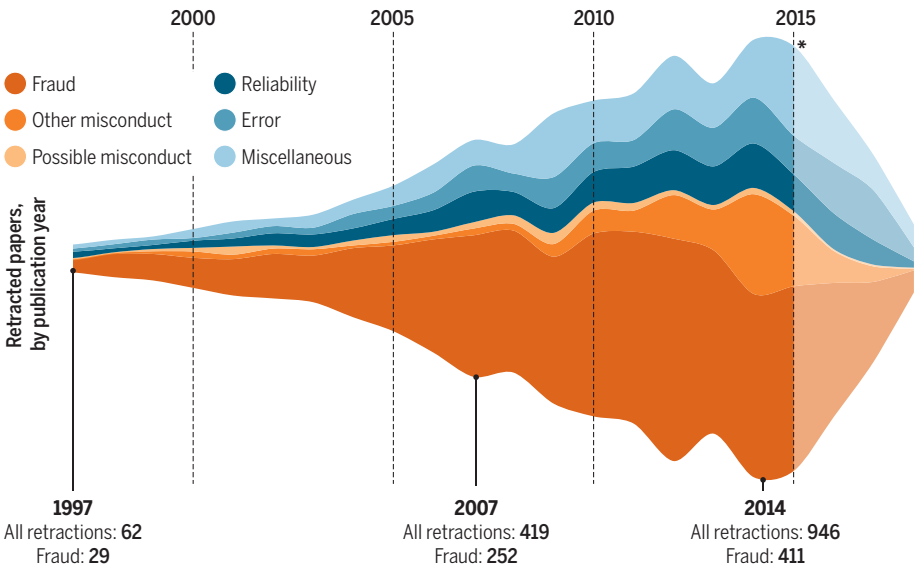
Science performed all data analyses for this report in consultation with Retraction Watch, using data drawn from the database on 30 August.

In analyzing the data, we made several choices. For example, we excluded nearly 7500 retracted conference paper abstracts, which account for about 40% of all the retractions in the database. Also, we analyzed retractions according to when the original paper was published, not the year the retraction notice appeared. For more details about our analytic methods, go to www.scim.ag/RWmethodology.

DATA: RETRACTION WATCH AND NSF

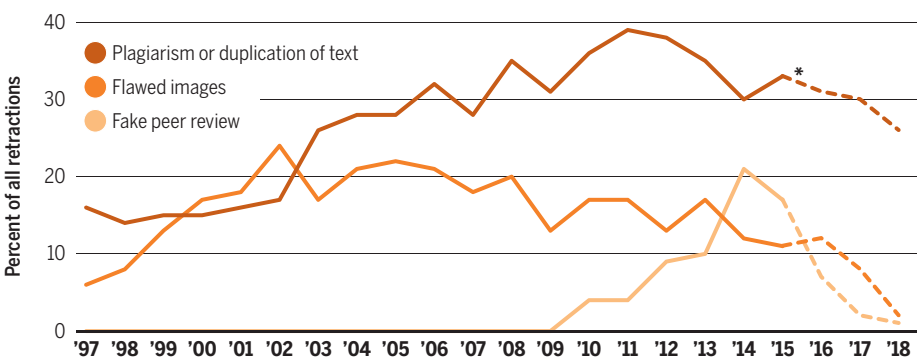
The burden of misconduct

The majority of retractions have involved scientific fraud (fabrication, falsification, and plagiarism) or other kinds of misconduct (such as fake peer review).



Changing infractions

The proportion of retractions involving plagiarism of text—stealing someone else's or duplicating one's own—has risen; one cause appears to be the introduction in 2004 of iThenticate, an internet-based plagiarism detection service. Fake peer reviews occur when authors give journals email addresses that they control, allowing them to review their own manuscripts. Flawed images include instances of intentional manipulation and of error.



*Retraction numbers appear to decline after 2015, but are almost certainly incomplete; journals typically take several years to retract papers. More information here: www.scim.ag/RWmethodology

Steneck says. Many journals now use software to detect plagiarism in manuscripts before publication, which can avoid retractions after.

But evidence suggests more editors should step up.

A disturbingly large portion of papers—about 2%—contain "problematic" scientific images that experts readily identified as deliberately manipulated, according to a study of 20,000 papers published in *mBio* in 2016 by Elisabeth Bik of Stanford University in Palo Alto, California, and colleagues. What's more, our analysis showed that most of the 12,000 journals recorded in Clarivate's widely used Web of Science database of scientific articles have not reported a single retraction since 2003.

Relatively few authors are responsible for a disproportionate number of retractions.

Just 500 of more than 30,000 authors named in the retraction database (which includes co-authors) account for about one-quarter of the 10,500 retractions we analyzed (see graphic, p. 394). One hundred of those authors have 13 or more retractions each. Those withdrawals are usually the result of deliberate misconduct, not errors.

Nations with smaller scientific communities appear to have a bigger problem with retractions.

Retraction rates differ by country, and variations can reflect idiosyncratic factors, such as a particularly active group of whistleblowers publicizing suspect papers (see sidebar and graphic, p. 395). Such confounding factors make comparing retraction rates across countries harder, Fanelli says. But generally, authors working in countries that have developed policies and institutions for handling and enforcing rules against research misconduct tend to have fewer retractions, he and his colleagues reported in *PLOS ONE* in 2015.

A retraction does not always signal scientific misbehavior.

Many scientists and members of the public tend to assume a retraction means a researcher has committed research misconduct. But the Retraction Watch data suggest that impression can be misleading.

The database includes a detailed taxonomy of reasons for retractions, taken from retraction notices (although a minority of notices don't specify the reason for withdrawal). Overall, nearly 40% of retraction notices did not mention fraud or other kinds of misconduct. Instead, the papers were retracted because of errors, problems with reproducibility, and other issues.

About half of all retractions do appear to have involved fabrication, falsification, or plagiarism—behaviors that fall within the U.S. government's definition of scientific misconduct. Behaviors widely understood within science to be dishonest and unethical, but which fall outside the U.S. misconduct definition, seem to account for another 10%. Those behaviors include forged authorship, fake peer reviews, and failure to obtain approval from institutional review boards for research on human subjects or animals. (Such retractions have increased as a share of all retractions, and some experts argue the United States should expand its definition of scientific misconduct to cover those behaviors.)

Determining exactly why a paper was withdrawn can be challenging. About 2% of retraction notices, for example, give a vague reason that suggests misconduct, such as an "ethical violation by the author." In some of those cases, authors worried about damage to their reputations—and perhaps even the threat of libel lawsuits—have persuaded editors to keep the language vague. Other notices are fudged: They state a specific reason, such as lack of review board oversight, but Retraction Watch later independently discovered that investigators had actually determined the paper to be fraudulent.

CREDITS: (GRAPHIC) J. YOU/*SCIENCE*; (DATA) RETRACTION WATCH

Ironically, the stigma associated with retraction may make the literature harder to clean up.

Because a retraction is often considered an indication of wrongdoing, many researchers are understandably sensitive when one of their papers is questioned. That stigma, however, might be leading to practices that undermine efforts to protect the integrity of the scientific literature.

Journal editors may hesitate to hand down the death penalty—even when it's justified. For instance, some papers that once might have been retracted for an honest error or problematic practices are now being "corrected" instead, says Hilda Bastian, who formerly consulted on the U.S. National Library of Medicine's PubMed database and is now pursuing a doctorate in health science at Bond University in Gold Coast, Australia. (The Retraction Watch database lists some corrections but does not comprehensively track them.) The correction notices can often leave readers wondering what to think. "It's hard to work out—are you retracting the article or not?" Bastian says.

COPE has issued guidelines to clarify when a paper should be corrected, when it should be retracted, and what details the notices should provide. But editors must still make case-by-case judgments, says Chris Graf, the group's co-chair and director of research integrity and publishing ethics at Wiley, the scientific publisher based in Hoboken, New Jersey.

A concerted effort to reduce the stigma associated with retractions could allow editors to make better decisions. "We need to be pretty clear that a retraction in the published literature is not the equivalent of, or a finding of, research misconduct," Graf says. "It is to serve a [different] purpose, which is to correct the published record."

One helpful reform, some commentators say, would be for journals to follow a standardized nomenclature that would give more details in retraction and correction notices. The notices should specify the nature of a paper's problems and who was responsible—the authors or the journal itself. Reserving the fraught term "retraction" for papers involving intentional misconduct and devising alternatives for other problems might also prompt more authors to step forward and flag their papers that contain errors, some experts posit.

SUCH DISCUSSIONS UNDERSCORE how far the dialogue around retractions has advanced since those disturbing headlines from nearly a decade ago. And although the Retraction Watch database has brought new data to the discussions, it also serves as a reminder of how much researchers still don't understand about the prevalence, causes, and impacts of retractions. Data gaps mean "you have to take the entire literature [on retractions] with a grain of salt," Bastian says. "Nobody knows what all the retracted articles are. The publishers don't make that easy."

Bastian is incredulous that Oransky's and Marcus's "passion project" is, so far, the most comprehensive source of information about a key issue in scientific publishing. A database of retractions "is a really serious and necessary piece of infrastructure," she says. But the lack of long-term funding for such efforts means that infrastructure is "fragile, and it shouldn't be."

Ferric Fang, a clinical microbiologist at the University of Washington in Seattle who has studied retractions, says he hopes people will use the new database "to look more closely at how science works, when it doesn't work right, and how it can work better." And he believes transparent reporting of retractions can only help make science stronger. "We learn," he says, "from our mistakes." ■

One publisher, more than 7000 retractions

By **Alison McCook**, *Retraction Watch*

Some 40% of the retractions in the Retraction Watch database have a single curious origin. Over the past decade, one publisher—the Institute of Electrical and Electronics Engineers (IEEE) in New York City—has quietly retracted thousands of conference abstracts.

Most of the abstracts are from IEEE conferences that took place between 2009 and 2011. The 2011 International Conference on E-Business and E-Government alone resulted in retractions of more than 1200 abstracts. In all, IEEE has retracted more than 7300 such abstracts. Most of the authors are based in China, and their papers covered topics as diverse as physical sciences, business, technology, and social sciences.

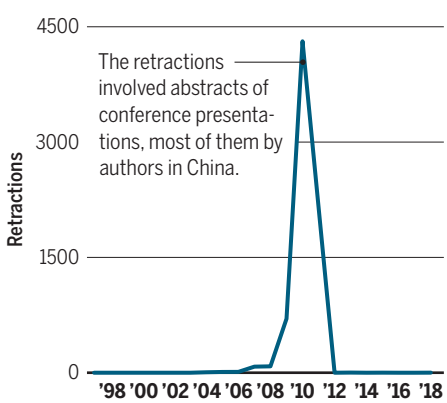
Many of the retraction notices offer few specifics about the reason. For example, the notice for retracting "The Study on Simulating Binaural Room Impulse Response" says simply: "After careful and considered review of the content of this paper by a duly constituted expert committee, this paper has been found to be in violation of IEEE's Publication Principles."

So what happened? IEEE hasn't given many details. The group, which sponsors more than 1700 conferences each year, requires peer review of all abstracts and papers before publication. But several years ago, in its decades-old catalog of abstracts, IEEE staff started to notice thousands of summaries that "did not meet our guidelines," according to a spokesperson. The spokesperson wouldn't disclose how they noticed the issue, "for reasons of operational integrity."

The episode may reflect the more rapid and less intensive form of peer review that conference submissions often undergo compared with papers submitted to traditional journals, says computer scientist Lior Pachter of the California Institute of Technology in Pasadena. The accelerated timetable "allows for quick turnaround for ideas and quick sharing," he says, but it also can mean that mistakes slip through.

To prevent future mass retractions, IEEE says it has formed a committee of staff and volunteer experts to serve as "gatekeepers" for conference materials and provide an additional level of quality control. That sounds like a good step, Pachter says. Researchers in quick-moving fields such as computer science "know and have known for a long time that many [conference] papers are problematic," he notes. And "people don't want to have garbage in their conferences."

Institute of Electrical and Electronics Engineers's spike in retractions



More information here: www.scim.ag/RWmethodology

CREDITS: (GRAPHIC) J. YOU/*SCIENCE*; (DATA) RETRACTION WATCH

A scientist's fraudulent studies put patients at risk

By **Adam Marcus**, *Retraction Watch*

In biomedical science, most papers that lead to retractions don't threaten anyone's life. But medical studies published by a once-prominent anesthesiologist offer a troubling exception, a story of how flawed and fraudulent papers can put patients in danger.

By 2010, Joachim Boldt was a research leader at Klinikum Ludwigshafen, an academic teaching hospital in Germany. That year, a sharp-eyed reader noticed a suspicious figure in one of Boldt's 2009 publications, in the journal *Anesthesia & Analgesia*. A later investigation revealed that Boldt had likely fabricated data, ignored ethics rules, and committed other kinds of misconduct in 98 articles he published with co-authors. All but two are now retracted.

Many of those studies had supported the effectiveness of intravenous solutions containing hydroxyethyl starch, or hetastarch, which doctors use to stabilize the blood pressure of patients during and after surgery or trauma. Although hetastarch and related products have been in widespread use since the 1960s, they are controversial. Study findings have suggested such products can have side effects including kidney damage and death. But Boldt's research appeared to show that a particular form, containing synthetic molecules called colloids, was safe.

That conclusion is now in doubt. In 2011, after the scandal, a group of medical societies in the United Kingdom withdrew their influential guidelines on intravenous fluid therapy—which endorsed colloids—because they included references to four of Boldt's tainted papers.

Joachim Stumpp, Boldt's former boss at Klinikum Ludwigshafen, told *Science* and Retraction Watch that, as far as he knows, investigators have found "no reported [cases] of serious impairment or injury of any patient" treated by Boldt (whose current whereabouts and work *Science* could not determine). But many other patients around the world likely were touched by the fraud, says Christian Wiedermann, an intensive care specialist at the Private University for Health Sciences, Medical Informatics and Technology in Hall in Tyrol, Austria, who has written several papers about the scandal. Although proving that particular patients suffered or died because of the misconduct would likely require expensive, large-scale studies, he says, "Logic dictates that, shamefully, patient harm must undoubtedly have resulted from Boldt's actions."

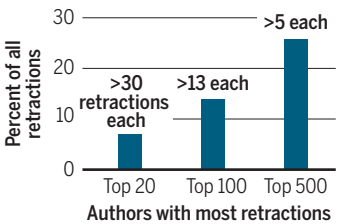
Researchers from the University of Manitoba in Winnipeg, Canada, have tried to assess how Boldt's misconduct might have distorted the scientific literature. Their analysis, published in 2013 in *The Journal of the American Medical Association*, examined 38 published articles, including seven by Boldt, comparing the use of fluids containing hetastarch with three other types of volume expanders. Taken together, the findings of the studies—which involved nearly 11,000 critically ill patients—suggested hetastarch solutions were as safe as the other fluids. But when the researchers excluded the 590 patients in Boldt's studies on hetastarch, a darker picture emerged: The fluids posed modest but statistically significantly greater risks of kidney damage and death.

The future of hetastarch treatments remains uncertain. This year, the European Medicines Agency, which has already curtailed the use of the products, proposed banning them outright.

German authorities reportedly considered bringing criminal charges against Boldt but have not.

A few authors, many retractions

Out of more than 30,000 authors in Retraction Watch's database, a handful accounts for a disproportionately large share of all retractions.



Top 10

Yoshitaka Fujii, Japan	169
Joachim Boldt, Germany	96
Diederik Stapel, Netherlands	58
Chen-yuan Peter Chen, Taiwan	43
Yoshihiro Sato, Japan	43
Hua Zhong, China	41
Shigeaki Kato, Japan	39
James Hunton, United States	36
Hyung-in Moon, South Korea	35
Jan Hendrik Schön, United States	32

More information here: www.scim.ag/RWmethodology

Fallout for co-authors

By **Alison McCook**, *Retraction Watch*

In 2011, chemist Bernhard Biersack, a postdoctoral fellow at the University of Bayreuth in Germany, struck a promising deal to collaborate with a well-funded cancer scientist based in the United States. They launched a multiyear partnership that Biersack says produced 12 journal articles, including seven that reported original research.

But the seemingly productive alliance soon became Biersack's worst nightmare: He had unknowingly signed on with a scientist who would be found guilty of scientific misconduct. That researcher, Fazlul Sarkar, formerly of Wayne State University in Detroit, Michigan, has now had more than 30 of his papers retracted.

You don't have to look hard for other cases of collaborators ensnared in scandals over fraudulent publications. In one high-profile case, social psychologist Diederik Stapel's tendency to make

CREDITS: (GRAPHIC) J. YOU/*SCIENCE*; (DATA) RETRACTION WATCH; (ILLUSTRATION) DAVIDE BONAZZI/SALZMANART

Volunteer watchdogs pushed a small country up the rankings

By **Ivan Oransky**, *Retraction Watch*

Drumroll, please: The countries that top our rankings of most retractions by nation are … Iran and Romania.

Why? They're not among the world's leaders in the absolute number of retractions—that dubious honor goes to the United States and China.

But ranking countries that way can be misleading. The United States and China fund many researchers who together publish many papers, which in turn can increase the number of papers that must be retracted.

Instead, *Science* and Retraction Watch created two measures that allow consistent comparisons across countries. The first is retractions per dollar of national research funding from 2003 to 2016, which is a proxy for the size of a nation's scientific establishment. The second is retractions per paper published.

By the funding measure, Romania takes the top spot. (The United States falls to 34th, and China to 14th.) But the story doesn't end there: Romania's leading rate of retractions per research dollar probably reflects the outsize effect of some dogged watchdogs—a small band of researchers who have been politely but firmly contacting journals to point out suspected plagiarism by Romanian authors. That activism has led to dozens of retracted papers.

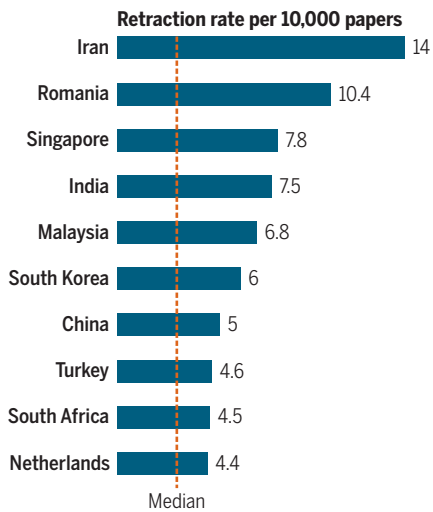
The effort was launched in 2013 by Stefan Hobai of the University of Medicine and Pharmacy in Târgu Mureş, Romania, who dubbed it, with no intentional irony, the Project dedicated to arrest of the name decline of the Romanian achievement (PANDORA) in biomedical publishing. Hobai tells *Science* and Retraction Watch that he acted because the editors of *Acta Medica Marisiensis*—published by the same university Hobai works for—ignored 17 messages in which he reported articles suspected of plagiarism.

Since then, PANDORA's few members, who other than Hobai have remained anonymous, have posted allegations on two blogs, including side-by-side comparisons of similar text. The editor of *Acta Medica Marisiensis* has questioned Hobai's motives and called his allegations "vicious." Editors at three other publications have acted on PANDORA's allegations, but not always in ideal fashion: They often simply removed plagiarized papers from their websites with no notification or reason, leaving no trace that the paper ever existed. That practice runs contrary to guidelines issued by the Committee on Publication Ethics, an international group that advises journal editors.

PANDORA is just one example of bands of researchers taking it upon themselves to clean up the literature. Some are subject specific, and many, such as PubPeer, are international. Others—such as VroniPlag, which began as a way to crowdsource suspected plagiarism in theses in Germany, and PANDORA—are country specific.

By the second measure—retractions per paper published—Iran tops the leaderboard and Romania drops to second (among countries that published at least 100,000 papers from 2003 to 2016; see graphic, left). Iran's position may reflect several high-profile scandals involving fake peer review. But this analysis may overstate Iran's retraction rate. That's because the incidence was calculated using a tally of published papers developed by the U.S. National Science Foundation. That count includes only papers published in English. If it also included papers published in Farsi—Iran's national language—the rate could change.

Countries with the highest retraction rates



More information here: www.scim.ag/RWmethodology

CREDITS: (GRAPHIC) J. YOU/*SCIENCE*; (DATA) RETRACTION WATCH AND NSF

up entire experiments led to dozens of retracted papers, most of which included junior collaborators.

So how do such disasters affect their careers? The short answer is: It depends.

Some collaborators face a frustrating struggle to clear their names. Thomas Hall, a professor of accounting at the University of Texas in Arlington, has repeatedly implored the publisher of a 2002 paper he co-wrote to reconsider its 2015 decision to retract it. Hall says the paper was withdrawn simply because another author, James Hunton, was found guilty of sweeping misconduct. Hall argues the results reported in their paper are valid and have been supported by later research. (The publisher, the American Accounting Association, didn't respond to requests for comment.)

In other cases, co-authors escape relatively unscathed. Biersack, for instance, has not been a co-author on any of Sarkar's retracted papers. Still, when Biersack learned about the misconduct, he was worried: Sarkar had contributed data and wording to some of his publications. So "I checked my papers with him again," he says. "I could not find mistakes."

Biersack remains a postdoc in Bayreuth, working on a temporary contract. He says he has seen no signs that his collaboration with Sarkar has held him back; no referees have mentioned it in their reviews of his work, for example.

His experience is consistent with findings reported by Joshua Krieger and colleagues at Harvard Business School in Boston in 2017. They showed that more prominent authors of papers retracted for fraud or misconduct often face greater penalties—in the form of fewer citations to their previous work—than do less prominent authors. But a different, 2013 study found that when it's not obvious who on a research team was to blame for a retraction, the less prominent co-authors experience larger declines in citations, reported Ginger Zhe Jin of the University of Maryland in College Park and colleagues.

To avoid possible career damage, Krieger suggests scientists build a portfolio of papers that includes ones written with different co-authors, which can help make a researcher "less sensitive to the discrediting of any one paper or researcher." But even if a co-author is hit with retractions, Biersack says, "it does not mean the end of your career."

EVOLUTION

How fish get their stripes—again and again

Cichlid fish emerge as new models for pattern development and evolution

By **Hugo F. Gante**

Conspicuous or otherwise elaborate color patterns sported by fishes, which constitute half of all living vertebrates, are among the most fascinating, varied, and intricate of nature. The importance of coloration for fishes can be appreciated from its many roles in camouflage and communication. The molecular and cellular mechanisms underlying color pattern development and evolution in teleost fish have received considerable attention, mostly through work in the zebrafish, *Danio rerio*, and its relatives. In these species, adult horizontal stripes, vertical bars, and spots along the body are formed by the timely expression of relevant genes and the orderly migration and interactions between three cell types originating from embryonic neural crest precursors: dark melanophores, yellow and orange xanthophores, and iridescent iridophores. These cells migrate to their final destination in the hypodermis and interact with each other at short and long ranges, and with their tissue environment, to produce color patterns predicted by Turing reaction–diffusion models, in which periodic patterns are produced by the interaction of an activator and an inhibitor (*1*–*6*). On page 457 of this issue, Kratochwil *et al.* (*7*) show that the degree of expression of the newly identified color pattern gene, agouti-related peptide 2 (*agrp2*), in the skin acts as a molecular switch mechanism controlling the presence or absence of horizontal melanic stripes in African cichlid fish.

Cichlids from the East African Great Lakes—Victoria, Tanganyika, and Malawi—have the highest sustained rates of speciation among vertebrates. Therefore, they offer perhaps the most diverse array of phenotypes for studying color pattern development and evolution in vertebrates. Kratochwil *et al.* used a combination of crosses between striped and nonstriped species pairs, fine-mapping of F_2 hybrid offspring and populations, gene expression assays, and functional experiments to narrow down the causative region to a 1.1-kb portion of the first intron that cis regulates expression of *agrp2*. These data indicate that

Department of Biology, The Center for Genomic Advocacy (TCGA), College of Arts and Sciences, Indiana State University, Terre Haute, IN 47809, USA. Email: hugo.gante@indstate.edu

PHOTO: KEVIN BAUMAN

Re-evolution (gain and loss) of horizontal stripes in cichlid fish from Lakes Victoria, Tanganyika, and Malawi is controlled by labile regulatory switches of *agrp2* expression.

PHOTOS: (TOP TO BOTTOM) FRANK TEIGLER/HIPPOCAMPUS BILDARCHIV; AD KONINGS

higher expression of *agrp2* represses horizontal melanic stripes in African cichlids.

Antimelanogenic *Agouti* genes have been implicated in color patterning in tetrapods through embryonic expression and establishment of a prepattern (*8*, *9*). Although *agrp2* expression differences in the skin of striped and nonstriped cichlids have been identified in adults, there are no obvious local differences that could explain the coordinates of stripe location. Indeed, *agrp2* expression levels are similar in melanic and nonmelanic skin (*7*). Thus, it is possible that an embryonic prepattern mechanism, conserved across vertebrates and involving any of the *Agouti* ohnologs (genes that arise from whole-genome duplication), also exists in cichlids. Alternatively, melanophore migration and patterning might be following particular anatomical landmarks, defining a boundary and followed by more localized interactions with other pigment cell types and the extracellular environment, to shape stripes in their correct location, possibly following Turing principles. Interestingly, melanic bars seem unaffected by the expression patterns of *agrp2* in the adult skin. This suggests the existence of different melanophore lineages, each with distinct bar or stripe fates, such as melanophores of embryonic or postembryonic origin in *Danio* species (*10*). Ablation of *agrp2* expression with CRISPR-Cas9 reconstitutes a midlateral stripe in a nonstriped cichlid but not a dorsolateral stripe. This points to additional genetic or cellular factors contributing to stripe phenotypes in cichlids.

Owing to their high species and phenotypic diversity, yet close phylogenetic relationship and availability of genome resources (*11*), the cichlids enable testing for molecular mechanisms involved in parallel or convergent evolution like no other vertebrate model. Stripes are lost and gained multiple times according to high or low expression of *agrp2* in the skin in different cichlid radiations from Lakes Victoria, Tanganyika, and Malawi (*7*). Interestingly, regulation of *agrp2* expression is achieved by different, rather than the same, genetic polymorphisms. These findings highlight how cis regulation might be a primary mechanism for generating phenotypic diversity from the same genes without requiring functional differences in protein sequence, particularly if those genes have a cascading effect on phenotype based on timing and location of their expression (*2*, *5*, *8*, *12*, *13*). Hybridization and recombination-mediated reshuffling of developmentally important cis-regulatory regions is a plausible mechanism to achieve high phenotypic diversity in a short period of time, as in cichlids from Lakes Victoria and Malawi, especially given the unexpectedly low mutation rates (*14*, *15*).

Much of the variation in the thousands of cichlid species is in the form of exacerbated color patterns and craniofacial phenotypes, which originate from a vertebrate evolutionary innovation—the neural crest. It seems that cichlids have harnessed this embryonic tissue to become one of the most diverse groups of vertebrates alive, and we now have better tools to investigate that. The study of Kratochwil *et al.* illustrates how advances in genetic engineering, driven by CRISPR-Cas9, are bringing classical models in ecology, evolution, and behavior to the molecular developmental biology laboratory. A new era has begun in which it is possible to demonstrate not only how cichlids get their stripes and how that might differ from mechanisms shaping zebrafish and tetrapod stripes but also how the seemingly independent stripe and bar modules of cichlids are regulated and how cell-cell interactions are coordinated during color pattern formation. It will finally be possible to answer long-standing questions about how the astonishing phenotypes of cichlids (head, body, and fin coloration; craniofacial shape; oral and pharyngeal dentition; fin shape; social and reproductive behaviors; breeding systems; sex, and even extended phenotypes like mating structures) are produced and often replicated throughout evolutionary history. ■

REFERENCES

1. S. K. McMenamin *et al.*, *Science* **345**, 1358 (2014).
2. L. B. Patterson *et al.*, *Nat. Commun.* **5**, 5299 (2014).
3. B. Walderich *et al.*, *Nat. Commun.* **7**, 11462 (2016).
4. A. Eskova *et al.*, *Development* **144**, 2059 (2017).
5. J. E. Spiewak *et al.*, *PLOS Genet.* **14**, e1007538 (2018).
6. M. Watanabe, S. Kondo, *Trends Genet.* **31**, 88 (2015).
7. C. F. Kratochwil *et al.*, *Science* **362**, 457 (2018).
8. M. Manceau *et al.*, *Science* **331**, 1062 (2011).
9. N. Haupaix *et al.*, *Science* **361**, eaar4777 (2018).
10. I. K. Quigley *et al.*, *Development* **131**, 6053 (2004).
11. D. Brawand *et al.*, *Nature* **513**, 375 (2014).
12. M. E. Santos *et al.*, *Nat. Commun.* **5**, 5149 (2014).
13. R. B. Roberts *et al.*, *Science* **326**, 998 (2009).
14. J. I. Meier *et al.*, *Nat. Commun.* **8**, 14363 (2017).
15. W. Salzburger, *Nat. Rev. Genet.* 10.1038/s41576-018-0043-9 (2018).

10.1126/science.aav3373

NEUROPHYSIOLOGY

The mechanotransduction of blood pressure

Two stretch-activated channels in arteries could be targets for some types of heart failure

By **Heimo Ehmke**[1,2]

The arterial baroreceptor reflex is the most important mechanism for minimizing short-term arterial blood pressure fluctuations (*1*). In situations of sudden blood pressure drops, the baroreceptor reflex accelerates heart rate, increases cardiac contractility, and induces vasoconstriction. Conversely, sudden increases in blood pressure trigger the opposite response. The autonomic nervous system mediates these physiological reactions. Patients with baroreceptor reflex malfunctions typically suffer from orthostatic hypotension, a severe decrease in blood pressure that occurs when standing up, which leads to dizziness or even fainting (*2*). However, how blood pressure changes are converted into electrical signals for neurotransmission has remained a puzzle. On page 464 of this issue, Zeng *et al.* (*3*) show that the mechanosensitive ion channels PIEZO1 and PIEZO2 are transducers of blood pressure in the sensory neurons of the autonomic nervous system that trigger the baroreceptor reflex. Identifying the molecular players of this response may help clarify the role of arterial baroreceptors in maintaining normotension and help develop new drugs for the treatment of heart failure.

The terminals of sensory neurons branch into thin fibers within the vascular wall of the aortic arch, which is directly downstream of the heart, and the carotid sinus in the neck, where the carotid artery divides before entering the brain. Unusually, the vascular wall at these sites contains little collagen, which greatly increases its elasticity. According to Laplace's law, even very small changes in transmural pressure cause large variations in vascular wall tension, facilitating mechanotransduction. This specific anatomy and the fact that pressure is the physiological signal to be transduced led to the assumption that the primary transducer of the baroreceptor reflex is a mechanosensitive ion channel. However, proof was lacking, owing to the inaccessibility of these thin nerve terminals within the vascular wall to electrophysiological measurements.

An alternative strategy to identify the baroreceptor reflex transducer is to inhibit the suspected mechanotransducer directly and to subsequently investigate the secondary consequences of baroreceptor reflex activation. This approach was applied previously (*4*–*7*), but a baroreceptor reflex response remained after genetic or pharmacological inhibition of the suspected mechanotransducer.

PIEZO channels play a role in mechanosensitive processes such as the perception of touch stimuli (*8*), lung inflation (*9*), and shear force on endothelial cells (*10*) in mammals. So far, only two isoforms of PIEZO channels are known, PIEZO1 and PIEZO2. They exhibit considerable differences in expression: PIEZO1 is expressed predominantly in non-neuronal tissues, and PIEZO2 is particularly found in sensory neurons (*11*). Therefore, PIEZO2 is an attractive candidate for the mechanotransduction of blood pressure in baroreceptor afferents. Surprisingly, Zeng *et al.* found in mice that both channels were expressed in the sensory neurons of the nodose and petrosal ganglion, which transmit the signal from the transduction site to the medullary cardiovascular center.

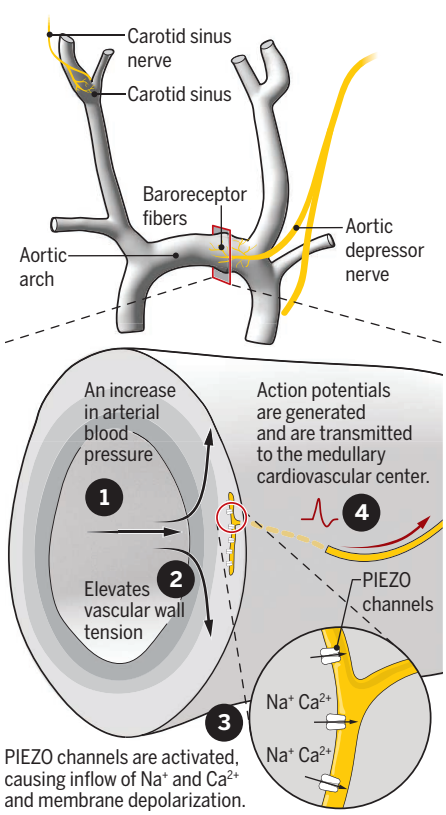
Furthermore, Zeng *et al.* show that only the combined ablation of both PIEZO channels suppressed baroreflex-mediated heart rate changes and eliminated activation of the sensory neurons mediating the baroreceptor reflex induced by a rapid increase in blood pressure. In accordance with the loss of the blood pressure–regulating function of the baroreceptor reflex, blood pressure instability was significantly increased in these animals. These results demonstrate that PIEZO1 and PIEZO2 are necessary for the mechanotransduction of blood pressure.

To test whether activation of PIEZO channels is also sufficient to trigger a baroreceptor reflex response, Zeng *et al.* used an optogenetic mouse model in which sensory afferents of the vagal nerve that express the *Piezo2* gene additionally express a photosensitive ion channel protein (*9*). Light stimulation of the carotid sinus region as well as of fibers of the aortic depressor nerve led to heart rate deceleration. Together, these findings convincingly demonstrate that PIEZO1 and PIEZO2 function as molecular sensors of blood pressure in arterial baroreceptors (see the figure). In addition, they answer the longstanding question of whether mechanotransduction takes place in the terminals of the sensory nerves or in neighboring cells.

The findings of Zeng *et al.* lead the way to new research avenues. Mechanosensitive receptors are also found in the low-pressure system of pulmonary circulation (in the lungs) and in the atria of the heart. Their activity is mainly determined by the amount of central blood volume and atrial contraction. Signals originating from these unknown mechanoreceptors influence the release of the hormone vasopressin, an important regulator of water balance, via the Gauer-Henry reflex and of the heart rate via the Bainbridge reflex. It would be interesting to examine whether activation of PIEZO channels also triggers these reflexes from the low-pressure system.

Arterial baroreceptors

Arterial blood pressure is continuously monitored by PIEZO1- and PIEZO2-expressing sensory neurons, which innervate the aortic arch and the carotid sinus. PIEZO-mediated activation of sensory nerves is responsible for the baroreceptor reflex, which maintains blood pressure and heart function.



PIEZO channels are activated, causing inflow of Na^+ and Ca^{2+} and membrane depolarization.

GRAPHIC: N. DESAI/*SCIENCE*

[1]*Department of Cellular and Integrative Physiology, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany.* [2]*Deutsches Zentrum für Herz-Kreislaufforschung (German Centre for Cardiovascular Research), Hamburg-Kiel-Lübeck, Germany. Email: ehmke@uke.de*

Arterial baroreceptors are indisputably important for short-term blood pressure regulation. Whether they also exert a long-term inhibitory influence on efferent sympathetic nerve activity remains controversial. The neurogenic hypertension observed by Zeng *et al.* after ablation of *Piezo1* and *Piezo2* is in favor of such an inhibitory function and supports the increasingly expressed view that arterial baroreceptors are also involved in the long-term regulation of blood pressure (*12*). Nonetheless, further studies are warranted to clarify the extent to which loss of cardiopulmonary reflexes also contribute to the development of hypertension observed by Zeng *et al.* The identification of PIEZO channels as blood pressure sensors opens fascinating possibilities for developing new genetic models for studying the physiological and pathophysiological role of baroreceptors in blood pressure regulation.

The study of Zeng *et al.* could provide the basis for the development of new drugs that activate PIEZO channels in order to suppress excessive sympathetic nerve activity. This could be relevant for patients with chronic heart failure, who suffer from a pathophysiologically increased efferent sympathetic activity, representing an important trigger for episodes of acute deterioration of cardiac pump function. The development of PIEZO channel–activating drugs is likely to be a difficult task. In view of the large number of important functions that PIEZO channels exert, a multitude of potential side effects must be considered, such as apnea (interrupted breathing), which can be triggered by excessive activation of PIEZO2 (*9*), and cardiac arrest. The development of a partial PIEZO channel agonist may overcome this problem. Understanding the physiological and clinical consequences of carrying the recently discovered gain-of-function allele of *PIEZO1*, which may cause baroreflex hyperactivation and is present in one-third of the African population (*13*), could also provide important information on the validity and clinical relevance of PIEZO channel agonists. ■

REFERENCES AND NOTES

1. D. L. Eckberg, P. Sleight, *Human Baroreflexes in Health and Disease* (Clarendon Press, Oxford, 1992).
2. R. Freeman *et al.*, *J. Am. Coll. Cardiol.* **72**, 1294 (2018).
3. W.-Z. Zeng *et al.*, *Science* **362**, 464 (2018).
4. H. A. Drummond *et al.*, *Neuron* **21**, 1435 (1998).
5. Y. Lu *et al.*, *Neuron* **64**, 885 (2009).
6. H. Sun *et al.*, *J. Pharmacol. Exp. Ther.* **331**, 851 (2009).
7. O. C. Lau *et al.*, *Nat. Commun.* **7**, 11947 (2016).
8. S. S. Ranade *et al.*, *Nature* **516**, 121 (2014).
9. K. Nonomura *et al.*, *Nature* **541**, 176 (2017).
10. J. Li *et al.*, *Nature* **515**, 279 (2014).
11. S. E. Murthy *et al.*, *Nat. Rev. Mol. Cell. Biol.* **18**, 771 (2017).
12. T. E. Lohmeier, R. Iliescu, *Physiology (Bethesda)* **30**, 148 (2015).
13. S. Ma *et al.*, *Cell* **173**, 443 (2018).

ACKNOWLEDGMENTS

H.E. is a consultant for preclinical cardiovascular research at Bayer.

10.1126/science.aav3495

ELECTRON MICROSCOPY

Pushing the limits of electron ptychography

Subatomic probe sizes in convergent-beam electron microscopy offer new opportunities

By **Andrew R. Lupini, Mark P. Oxley, Sergei V. Kalinin**

Ptychographic imaging in transmission electron microscopy is based on the detection of a huge number of diffraction patterns generated as an electron-beam probe scans the sample. Unlike conventional diffraction experiments where information about the phase of the scattered beam is lost, by scanning the entire sample and recording the diffraction patterns, an inversion analysis can reconstruct the complex amplitude of the beam and ultimately the atomic positions. Probe beams continue to decrease in size and can be much smaller than the distance between scatterers. The diffraction spots also broaden and interfere with one another (see the figure). Analysis of this interference allows more information to be extracted, such as minute details of nuclei position and electronic structure that define physical properties and enable visualization of subtle details of electric fields, orbital orderings, ferroic behavior, and perhaps local spins. Ptychography and related techniques are poised to revolutionize the field of electron microscopy, in much the same way as they have in x-ray crystallography (*1*), and as aberration correction has achieved for imaging, provided that some key challenges can be met.

This opportunity extends the innovations that have occurred since the invention of transmission electron microscopy in the 1930s, which extended the concept of diffraction from x-ray studies of macroscopic crystals to nanometer-scale volumes. The development of scanning transmission electron microscopy (STEM) (*2*) and especially aberration correction in STEM (*3*) offered the potential of direct atomic imaging by an atomically sized focused beam. As a result, the past 15 years have seen explosive growth of STEM applications in condensed-matter physics and materials science. Notably, the availability of the Z-contrast imaging mode provides a relatively easy way to interpret images, where contrast maxima can be directly related to the position of atomic nuclei (*4*).

Despite these advances in direct imaging, fundamentally new opportunities arise by detecting the convergent-beam electron diffraction (CBED) pattern in STEM, also known as a "Ronchigram" after Vasco Ronchi, who pioneered a similar technique in light optics. In this case, the size of the electronic probe can be smaller than the distance between the atomic nuclei, a paradoxical situation from the conventional physics perspective because diffraction is usually thought of as a nonlocalized effect, depending only on the spacing between scatterers. The process of Ronchigram evolution is illustrated in the figure. The resulting pattern contains information on both the beam shape, and minute details of material structure (nuclei position) and internal electric and magnetic fields. Recording these patterns at every position in a conventional image generates an intriguing multidimensional dataset, generally referred to as four-dimensional (4D) STEM, that opens up several new routes for analysis.

> ***"Subatomic diffraction... imaging opens a window into the atomic-scale functionalities of solids...on a level of single structural units."***

Ptychography was first proposed by Hoppe in the 1960s, where, just like the closely related holography, it was originally devised as a route to overcome the resolution limit in the electron microscope arising from lens aberrations (*5*). For STEM, the potential of ptychographic imaging began to be realized by Rodenburg and multiple coworkers in the 1990s, but general applications had to wait until sufficiently powerful cameras, computers, and iterative algorithms for data analysis were developed and

Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. Email: arl1000@ornl.gov

refined (*6*–*8*). The potential applications of these imaging modes were even further broadened when Shibata (*9*), using a segmented detector, demonstrated that differential phase contrast in STEM can transfer information about local electric fields with atomic resolution.

Despite the demonstration of atomic resolution through these methods, aberration correction is still critical, and the highest sensitivity will still depend on optimizing the rest of the microscope (*10*). Given the recent upswing of multidimensional data acquisition in STEM prompted both by the availability of pixelated detectors and anticipation of coming physics breakthroughs, the challenges need to be considered if these imaging methods are to be widely adopted (*11*).

Data throughput and storage are key areas that must be considered. A useful estimate for the coherent beam current in a modern high-resolution STEM is a few (tens of) picoamperes. At 32 pA, equivalent to 200 million electrons per second, recording a probe position and scattering direction for every electron would generate ~1 GB of data per second. This volume corresponds well to modern camera technologies and, uniformly distributed over a 4-megapixel camera of about 50 frames per second, would give an average of ~1 electron per pixel. However, the first step of most processing workflows presently consists of binning the data down to a more manageable size, either by integrating over position or angle. The challenge will be to condense this massive flow into a compressed stream consisting of only the useful information at the detector side, and/or to develop the infrastructure capable of capturing and storing these data volumes. Although these problems are generally well addressed in high-energy physics, astronomy, and scattering communities, they still need to find their way into the STEM field.

Another consideration is that the probe shape, even one that is aberration corrected, still has residual nonround aberrations and might not be at the ideal position. Unlike conventional imaging, spatial incoherence cannot be added after the fact by a simple convolution but must be included using an incoherent sum of CBED patterns taken at slightly different positions. For low voltages, as often used for two-dimensional materials (*11, 12*), the effect of temporal incoherence will complicate the analysis. This consideration is extremely important, because the Ronchigram is determined both by the beam shape and structural distortions in the material, and hence quantification of the underpinning physics is possible only by accounting independently for the beam shape.

Even for ideal beam shapes, determining the structure of the specimen requires progressively more complex theoretical models. Simple iterative ptychography may be used for suitably thin specimens, where various approximations (such as assuming that the object only has a small effect on the electron wavefront) are reasonable, to determine the projected potential. Differential phase contrast allows the measurement of local electric fields, which also can be related to the scattering potential for suitably thin specimens (*9*). However, the quantum-mechanical inversion of multiple scattering associated with thicker specimens is extremely complex. This situation is further complicated by inelastic scattering (primarily plasmons), which further changes the CBED patterns (*13*). Detecting CBED patterns as a function of defocus or sample tilt is an approach to address this problem; however, the dimensionality of the data and associated challenges grow further.

Presently, a robust solution requires a well-defined experiment, with limits or knowledge required for specimen thickness, probe parameters, aberrations, incoherence, and other parameters. Detailed (forward-scattering) simulations of model structures require model structures for density functional theory, as well as an accurate method to convert the calculated electron densities to scattering potentials. With these in hand and incorporation of automated workflows to generate multiple datasets, the use of deep-learning inversion algorithms becomes feasible. The general case will need these sort of capabilities, as the real breakthroughs will be found where experiment turns out to be different from expected models.

Subatomic diffraction, or 4D STEM, imaging opens a window into the atomic-scale functionalities of solids—disorder, ferroic distortions, and field and orbital structure—on a level of single structural units. Rapid realization of this opportunity necessitates initiating interdisciplinary research efforts combining electron microscopy, data analytics and image analysis (*14*), and machine learning, with special attention to the mechanisms of image formation in aberration-corrected machines. Initiating discussion in this area can serve as a much-needed seed to nucleate this effort and establish its rapid growth. ■

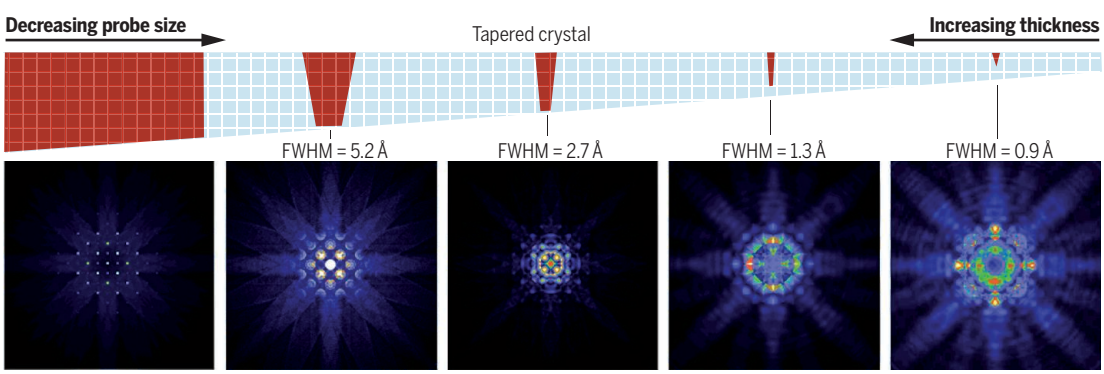
REFERENCES

1. J. Miao, T. Ishikawa, I. K. Robinson, M. M. Murnane, *Science* **348**, 530 (2015).
2. A. V. Crewe, *Microsc. Microanal.* **10**, 414 (2004).
3. O. L. Krivanek, N. Dellby, A. R. Lupini, *Ultramicroscopy* **78**, 1 (1999).
4. S. J. Pennycook, L. A. Boatner, *Nature* **336**, 565 (1988).
5. W. Hoppe, *Ultramicroscopy* **10**, 187 (1982).
6. J. M. Rodenburg, R. H. T. Bates, *Philos. Trans. R. Soc. Lond. A* **339**, 521 (1992).
7. A. M. Maiden, M. J. Humphry, M. C. Sarahan, B. Kraus, J. M. Rodenburg, *Ultramicroscopy* **120**, 64 (2012).
8. A. J. D'Alfonso *et al.*, *Phys. Rev. B* **89**, 064101 (2014).
9. N. Shibata *et al.*, *Nat. Phys.* **8**, 611 (2012).
10. A. Rosenauer, F. F. Krause, K. Müller, M. Schowalter, T. Mehrtens, *Phys. Rev. Lett.* **113**, 096101 (2014).
11. H. Yang *et al.*, *Nat. Commun.* **7**, 12532 (2016).
12. Y. Jiang *et al.*, *Nature* **559**, 343 (2018).
13. H. G. Brown *et al.*, *Ultramicroscopy* **182**, 169 (2017).
14. S. Jesse *et al.*, *Sci. Rep.* **6**, 26348 (2016).

10.1126/science.aau7620

Turn down the volume

Diffraction patterns change as the volume of material probed by an electron beam decreases. Subatomic diffraction patterns can be used to determine the minute details of local structure and scattering potential (and hence local electric fields) if the measurements are taken in multiple locations and the exact probe shape is known.



Large versus small probes
For large electron beams, the diffraction pattern typical of a crystal is observed (far left). Convergent beams have a full width at half maximum (FWHM) comparable to or less than unit-cell distance. The Fourier peaks broaden, giving rise to complex diffraction patterns.

GRAPHIC: N. CARY/*SCIENCE*; (IMAGES) MARK OXLEY

CANCER

The chromatin of cancer

Understanding tumorigenesis and inherited risk for cancer requires a multiomic approach

By **Jussi Taipale**[1,2,3]

Developments in modern genomics tools have led to rapid progress in our understanding of the genetic basis of cancer. Recent large-scale efforts have primarily focused on two types of analysis: mapping acquired somatic mutations by whole-exome and whole-genome sequencing (*1*, *2*), and identification of common inherited variants that increase cancer risk using genome-wide association studies (GWAS) (*3*). Despite the power of these technologies, we are still far from understanding how the variants and mutations found in individual tumors precisely drive the oncogenic process. A large number of genetic variants increase risk for cancer, but most explain only a very small fraction of the risk. Furthermore, although acquired somatic mutations are found in almost all tumors, most do not carry complete sets of mutations that, according to our present mechanistic understanding, would be sufficient to cause cancer. On page 420 of this issue, Corces *et al.* (*4*) show how a third type of genomics approach—functional genomic analyses of primary human tumors—can begin to bridge this gap in our mechanistic understanding of the tumorigenic process.

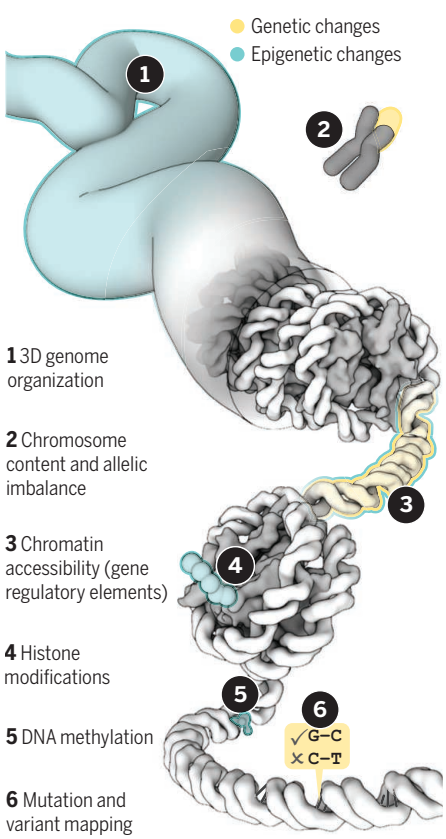
The authors analyzed chromatin accessibility using ATAC-seq (assay for transposase-accessible chromatin using sequencing) of 410 primary tumors representing 23 different types of human cancer. Analysis of chromatin accessibility measures stable binding of proteins to the genome; regions that are unbound are accessible to enzymes such as deoxyribonuclease I (DNase I) (*5*) or Tn5 transposase (*4*). The ATAC-seq method used by Corces *et al.* utilizes Tn5, which inserts a linker sequence to accessible DNA and cuts it, allowing highly efficient isolation and sequencing of the liberated fragments. Most of the human genome is relatively inaccessible because it is wound around histone proteins, forming nucleosomes, each of which contains 147 base pairs of DNA. In less than 1% of the genome, the histones are replaced by other proteins that regulate chromosome structure, or that function as transcription factors to direct gene expression. Tn5 can insert DNA linkers between such proteins; if the proteins are bound tightly, their binding position also leaves a "footprint" that is narrower than that formed by a nucleosome. DNA accessibility is known to correlate with the presence of active gene regulatory elements such as promoters and enhancers, and is thus commonly used as a proxy for gene regulation. Motif mining of the accessible regions and analysis of sequences under the footprints can then be used to infer which sequence-specific DNA binding proteins are bound to the accessible regions. The power of the approach of Corces *et al.* derives from the combination of deep sequencing that allows footprinting with the analysis of a large number of samples representing different types of cancer. Importantly, the samples used are sequenced for mutation mapping in The Cancer Genome Atlas (TCGA) project, facilitating comparative multiomic analyses between different data types.

The motif mining and footprinting analyses reveal many transcription factors that are strongly active in the different cancer types. For example, the authors detect androgen receptor in prostate cancer and microphthalmia-associated transcription factor (MITF) in melanoma, indicating that ATAC-seq can pinpoint known cancer type–specific transcription factors. The identification of accessible chromatin across multiple cancer types, together with detection of expressed genes by RNA sequencing, allows inference of DNA elements that may regulate gene expression (*5*). This analysis is based on correlation, but the authors also validate a subset of the potential enhancer-promoter links by targeting a repressor to the regulatory elements using CRISPR-Cas9 interference. Compared to RNA sequencing, the analysis of accessible chromatin also gives a more detailed "fingerprint" of the tissue, facilitating classification of tumors and analysis of their cellular composition. The chromatin accessibility data can also be used to locate elements that contain variants that may contribute to inherited cancer risk, and to identify somatic noncoding mutations that affect chromatin accessibility. Given the scale of the dataset and its multiomic character, there is great potential for new discoveries. Most of the individual findings reported by Corces *et al.* need further validation. However, the large number of interesting initial discoveries, such as the link between elements near the *MECOM* gene and adverse outcome in kidney cancer, highlights the value of the dataset as a reference and as a data-mining resource for future studies.

The analysis of chromatin accessibility in primary tumor cells also extends the known repertoire of potential gene regulatory elements. Of those identified by Corces *et al.*, 35% were not previously known. Many of the elements identified from primary tumors are likely to be important for normal developmental and homeostatic processes. However, there is good reason to suspect that at least some may not be so benign. Our genome is likely to encode a large number of potentially pathological transcription factor–DNA interactions (*6*, *7*). This is because cancer-causing mutations can directly affect transcription factor binding sites, leading to activation of normally silent regulatory elements (*8*). Mutations can also activate transcription factors

Sequence and function of the cancer genome

Multiomic mapping and comparison between genetic and epigenetic features are required for mechanistic understanding of cancer and for providing a "fingerprint" of the tumor. Multiomic analyses are likely to be important for cancer diagnosis and prediction of outcome, as well as for guiding treatment decisions and drug development.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

[1]*Department of Biochemistry, University of Cambridge, Cambridge CB2 1GA, UK.* [2]*Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden.* [3]*Genome-Scale Biology Program, University of Helsinki, Helsinki, Finland. Email: ajt208@cam.ac.uk*

that have low or no activity in the corresponding normal tissues, leading to activation of large sets of gene regulatory elements (*9*). In addition, many driver mutations affect chromatin modifiers or alter levels of CpG methylation of DNA, leading to destabilization of the entire chromatin regulatory system (*1*, *2*). It is difficult for evolutionary processes to remove large sets of potentially harmful elements from our genome, as each individual element has limited impact at the population level, and cancer generally affects individuals who are above reproductive age. Therefore, it is likely that elements that are specifically activated in cancer are present in our genome. Identification of such elements will facilitate improved diagnosis and prognosis, and also allow investigations of new therapeutic modalities to target oncogenic gene regulation.

The mapping of accessible chromatin landscapes is also important for the mechanistic understanding of tumorigenesis. It is known that altered activities of transcription factors and/or their binding regions drive the major forms of human cancer. Cancer can thus be considered a disease of gene expression, where a combination of mutations locks the gene regulatory network of a single cell into a state that drives unrestricted cell proliferation (*6*). Although mutations in some oncogenes and tumor-suppressor genes are commonly found across many forms of cancer, most driver genes are mutated in a more restricted set of tumors. Some of the differences in oncogene composition can be explained by differences in mutational mechanisms and proto-oncogene expression between the cell types of origin of the tumors. However, many oncogenes cannot transform fibroblastic cells in standard cell-based assays, suggesting that cell lineage–determining transcription factors collaborate in some way with oncogenes. The mechanisms of such collaboration are currently poorly understood, but given that lineage-determining factors commonly define chromatin states, it is likely that accessibility of chromatin at specific regulatory sites contributes to this process. An important contribution of the study by Corces *et al.* is the identification of candidate sets of such lineage-specific regulatory elements that are critical for the cancer phenotype.

Cancer genome sequencing efforts have revealed that a large number of genes can cause cancer. Because most of the driver genes are mutated infrequently, making mechanistic sense of the cancer genotype by straightforward genetic interaction analysis requires extremely large sample sizes. Combining genomic data with phenotypic information is thus an attractive alternative. Traditionally, there has been a disconnect between cancer genomics and large-scale efforts to map the functional genome. The Roadmap Epigenomics (*10*) and Genotype-Tissue Expression (GTEx) (*11*) projects primarily focus on normal tissues, whereas the main drive of ENCODE (*5*) is to identify functional genomic elements; although cancer cells are used as models in some of these projects, the cell lines used do not adequately represent major forms of human cancer. Previous epigenomic studies of cancer, in turn, have mainly focused on targeted DNA methylation analysis (*12*), transcription factor binding analyses in a few cell lines (*13*), or analysis of histone modifications in a particular type of cancer (*14*). In this context, the study by Corces *et al.* is particularly welcome because it paves the way toward a large-scale effort to map the functional genome of cancer cells. To understand how individual tumors form, it is necessary to map their genomic features such as germline variants, somatic mutations, chromosomal content, and allelic imbalance (*15*), together with functional genomic features such as genes that are essential for growth and survival, three-dimensional (3D) chromosome conformation, the DNA methylome, chromatin modification state, and accessible chromatin landscape (see the figure). Comparing cancer types to each other can yield interesting results but suffers from the disadvantage that all cancers share key phenotypic characteristics, such as unrestricted growth. A better comparison would be between cancers and their cell types of origin. However, the cell type of origin of many cancers is unknown, and many tumors are thought to originate from relatively rare cells (for example, stem or progenitor cells). Therefore, it will also be necessary to develop analytical methods that can detect genomic features from minor cell populations or from single cells. Without such multiomic maps at the cell-type level, it will be exceedingly difficult to move from genomics toward understanding the main drivers of the phenotype of individual tumors. Without such understanding, we may not be able to conquer cancer. ■

REFERENCES

1. L. A. Garraway, E. S. Lander, *Cell* **153**, 17 (2013).
2. B. Vogelstein *et al.*, *Science* **339**, 1546 (2013).
3. A. Sud *et al.*, *Nat. Rev. Cancer* **17**, 692 (2017).
4. M. R. Corces *et al.*, *Science* **362**, eaav1898 (2018).
5. R. E. Thurman *et al.*, *Nature* **489**, 75 (2012).
6. I. Sur, J. Taipale, *Nat. Rev. Cancer* **16**, 483 (2016).
7. J. Taipale, *EMBO J.* **37**, e96114 (2018).
8. M. R. Mansour *et al.*, *Science* **346**, 1373 (2014).
9. K. Gangwal *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10149 (2008).
10. P. Polak *et al.*, *Nature* **518**, 360 (2015).
11. M. Melé *et al.*, *Science* **348**, 660 (2015).
12. K. A. Hoadley *et al.*, *Cell* **173**, 291 (2018).
13. J. Yan *et al.*, *Cell* **154**, 801 (2013).
14. B. Akhtar-Zaidi *et al.*, *Science* **336**, 736 (2012).
15. K. Palin *et al.*, *Nat. Commun.* **9**, 3664 (2018).

10.1126/science.aav3494

EVOLUTION

Our shallow-water origins

Coastal habitats represent a cradle of diversification for early vertebrates

By **Catalina Pimiento**[1,2]

Vertebrates encompass all animals with a backbone, from fish to humans. How and when they evolved are questions that have been studied for centuries, revealing the origins and processes involved in anatomical innovations such as jaws, teeth, and paired appendages (*1*). A less explored, but equally important question is where they evolved. On page 460 of this issue, Sallan *et al.* (2) compile a new database of early occurrences (mid-Paleozoic, 490 to 360 million years ago) and site-specific environmental information to reconstruct vertebrate ancestral habitats. They report that all major vertebrate clades originated in restricted, shallow-water environments.

The environmental context of vertebrate evolution had remained a gap in our knowledge. Understanding the habitat constraints present when key traits evolved is necessary to answering fundamental questions in macroevolution such as the extent to which the environment can drive anatomical transformations. Reaching this level of understanding has been limited mostly by a lack of data in available compendia (*3*). The examination of primary data on early fish (e.g., from the mid-Paleozoic) revealed that their fossil record accumulated in shallow waters (*4*). However, it has been recognized that this might be an artifact of a poor fossil record; in other words, the habitats from where ancient fish have been recovered might reflect outcrop (the exposure of rocks)

"What was so special about the shallow-water habitats where vertebrates diversified?"

[1]*Department of Biosciences, Swansea University, Swansea, UK.* [2]*Smithsonian Tropical Research Institute, Panama. Email: cpimiento@swansea.ac.uk*

availability rather than true origins. Sallan *et al.* explicitly test this possibility and demonstrate that although fossils of early fish are mostly are found in rocks coming from depths between 60 and 200 m (*5*), the early diversification of vertebrates was restricted to shallower environments of less than 60 m of depth. Accordingly, the ancestral habitats of early fish are not a sampling artifact.

Importantly, Sallan *et al.* show that the use of shallow-water habitats as a cradle for diversification was robust and persistent over time. Similar to what has been found in benthic invertebrates (those living in or on the bottom sediments of the ocean floor) (*6*), vertebrates continued to originate in shallow waters even long after they had diversified, dispersed, or evolved anatomic innovations (e.g., jaws). Evolutionary shifts to deeper waters were far more difficult than to other nearshore environments, or to freshwater. Nevertheless, early fish managed to occupy different aquatic environments along the depth gradient. Interestingly, dispersal into habitats outside the cradle were not necessary to evolve new phenotypes. Instead, major body forms [benthic (those adapted to live on the bottom of the ocean) and pelagic (those adapted to live in the water column)] originated in shallow waters before expanding to new habitats.

What was so special about the shallow-water habitats where vertebrates diversified? The mid-Paleozoic nearshore environments were somewhat different from those of today. Seagrasses, mangroves, and modern coral reefs had yet to appear. Nevertheless, habitat-forming species such as stromatolites, sponges, and early corals were present (*5*). During this time, these habitats experienced fundamental evolutionary changes as the water column gradually filled with newly evolved nektonic forms (organisms able to freely swim) (*7*). Exploitation of the vertical habitat dimension was likely driven by competition in the saturated benthic zone, and by increased ocean productivity resulting from riverine influx when arborescent flora evolved on land (*8*). Although these transformations took place in all coastal habitats, most origination occurred in lagoon-like systems, and therefore in sheltered areas. Was the combination of heterogeneity, habitat structure, and protection the foundation for the diversification of major vertebrate clades?

Today, protected shallow-water ecosystems are not only biodiversity hotspots, but also serve as essential nurseries for fish (e.g., coral reefs, estuaries, and mangroves). These ecosystems offer physical structure, habitat heterogeneity, and trophic complexity, thus providing abundant food and refuge to marine fauna, as well as important services to humans (*9*). Nearshore systems have supported fish diversity for at least 66 million years (*10*). Sallan *et al.* not only extend this association to the very origins of vertebrates, but also highlight the role of shallow waters as a persistent cradle for their diversification. Nevertheless, just as these environments can support biodiversity, their reduction can also result in its loss. Between five and two million years ago, shallow-water habitats contracted as a result of dramatic sea-level oscillations, likely causing the extinction of a substantial number of marine vertebrates (*11*). Before these already-vulnerable organisms had time to recover, modern humans started degrading their (shallow-water) habitats by overexploiting their fauna and destroying the structure that provides the foundations of biodiversity (*12*). Sallan *et al.* show that without shallow-water ecosystems, vertebrates (humans included) would probably not have evolved. Worryingly, it is precisely these ecosystems that have been altered the most by human activities (*13*). ■

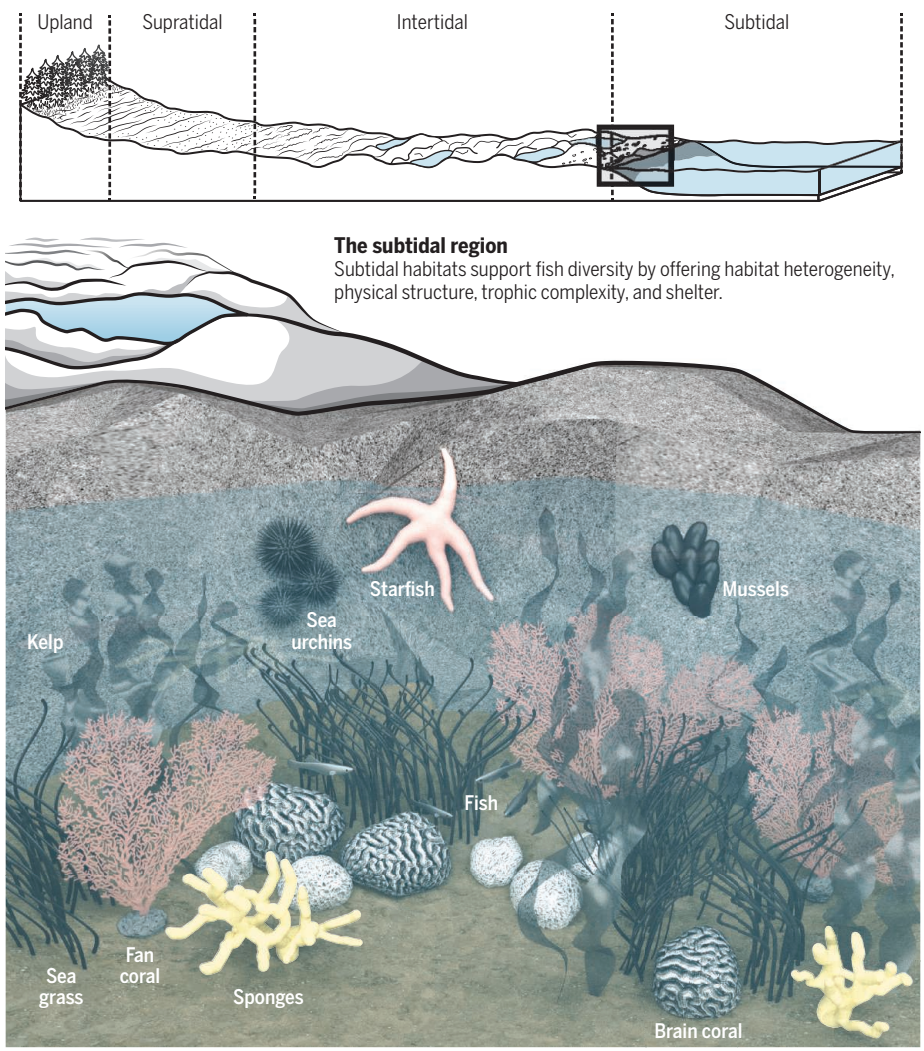
REFERENCES

1. M. D. Brazeau, M. Friedman, *Nature* **520**, 490 (2015).
2. L. Sallan *et al.*, *Science* **362**, 460 (2018).
3. M. Friedman, L. C. Sallan, *Palaeontology* **55**, 707 (2012).
4. R. S. Sansom, E. Randle, P. C. J. Donoghue, *Proc. R. Soc. B* **282**, 20142245 (2015).
5. C. E. Brett, A. J. Boucot, B. Jones, *Lethaia* **26**, 25 (1993).
6. D. Jablonski, J. J. Sepkoski Jr., D. J. Bottjer, P. M. Sheehan, *Science* **222**, 1123 (1983).
7. C. D. Whalen, D. E. G. Briggs, *Proc. R. Soc. B* **285**, 20180883 (2018).
8. C. Klug *et al.*, *Lethaia* **43**, 465 (2010).
9. E. B. Barbier *et al.*, *Ecol. Monogr.* **81**, 169 (2011).
10. D. R. Bellwood, C. H. R. Goatley, O. Bellwood, *Biol. Rev.* **92**, 878 (2017).
11. C. Pimiento *et al.*, *Nat. Ecol. Evol.* **1**, 1100 (2017).
12. H. K. Lotze *et al.*, *Science* **312**, 1806 (2006).
13. K. R. Jones *et al.*, *Curr. Biol.* **28**, 2506 (2018).

10.1126/science.aau8461

Nearshore environments are the cradle of vertebrate evolution

Shallow subtidal environments represent the ancestral habitats of early fish.



GRAPHIC: A. KITTERMAN/*SCIENCE*

POLICY FORUM

GLOBAL HEALTH

Threats to timely sharing of pathogen sequence data

The Nagoya Protocol may impose costs and delays

By Carolina dos S Ribeiro[1,2], Marion P. Koopmans[3], George B. Haringhuizen[1]

Pathogen genome sequence databases are taking over important functions of physical collections of microbial and viral cultures (biobanks), adding functionalities for worldwide rapid sharing of pathogen genetic resources in support of research and outbreak response (*1*). But biobanks and databases also have to respect the ownership and rights of the sample and data providers, including the sovereign right of states to decide on the use of their resources [as stated in the Convention on Biological Diversity (CBD) (*2*)]. Where domestic or international regulation in this regard is absent or unclear, the integrity of databases and biobanks can be threatened by divergent interpretations, potentially leading to perceived violation of globally agreed sovereignty rights. In particular, the impact of the Nagoya Protocol (NP) to the CBD on public health and infectious disease control is highly debated and focused now on whether genetic sequence data (GSD) fall within the scope of the NP, which thus far has concentrated on access to physical samples. With this question on the agenda of the upcoming CBD Conference of the Parties (17 to 29 November) (*3*), we explore possible adaptations of existing biobank frameworks to support efficient transfer of pathogen genetic resources (PGR) during public health emergencies.

Following the West African Ebola outbreak, the World Health Organization (WHO), in consultation with a broad group of stakeholders, concluded that "sharing relevant information before publication should become the global norm during public health emergencies…including pathogen sequences." If GSD would fall under the NP, the free sharing of sequences for surveillance and tracking of pathogens and outbreaks could become illegal (*4*).

[1]The Netherlands National Institute for Public Health and the Environment (RIVM), Center for infectious Disease Control, Bilthoven, Netherlands. [2]Athena Institute for Research on Innovation and Communication in Health and Life Sciences, Vrije Universiteit Amsterdam, Amsterdam, Netherlands. [3]Erasmus Medical Center, Department of Viroscience, Rotterdam, Netherlands. Email: ribeirocds@hotmail.com

TRANSFER OF GENETIC RESOURCES

The NP is an internationally binding treaty, adopted in 2010, to regulate access to genetic resources [thus far agreed for biological (physical) samples] and to ensure fair and equitable sharing of benefits arising from their use (*5*). The NP determines that governments can exercise sovereign rights over resources originating from their territory. To this end, national legislation has to be in place and made transparent, preferably by uploading into the CBD's Access and Benefit-Sharing (ABS) Clearing-House. Before users receive any sample from abroad, the NP requires them to check (i) whether the providing state has signed the NP; (ii) whether a law or regulation is available in the ABS Clearing-House, and if not; (iii) to check with the state's authorities if any national law prohibits the access and transfer of the material. Detailed recording of this procedure and correspondence can prove later due diligence if the necessary permit is not available.

With the NP entered into force in 2014, countries can state specific conditions in which the receiving party has to acquire two types of permissions. First, users must obtain prior informed consent (PIC) from the provider's country government. Second, the monetary and/or in-kind terms and conditions for utilization of the genetic resources [the so-called Mutually Agreed Terms (MAT)] need to be obtained from the provider's country government. Differentiation between commercial and noncommercial use is in the NP and for most countries quite essential when deciding on the type and value of ABS conditions in their MAT. The NP promotes the use of internationally recognized certificates and their uploading into the ABS Clearing-House for everyone to access, but other forms of agreements are possible [e.g., through letters or a contractual material transfer agreement (MTA)].

This responsibility of assuring that any shipment of samples is compliant to the NP by obtaining and transferring acceptable documentation presents a substantial burden of which many in the research and public health community are not well aware. For biobanks, where huge amounts of items are stored for multilateral sharing, using standard NP certificates is probably the most workable way of documenting compliance.

DEFINITION, IMPLEMENTATION

A core term in the NP is the "utilization of genetic resources," defined as "to conduct research and development (R&D) on the genetic and/or biochemical composition of genetic resources, including through the application of biotechnology" (*5*). But when specific activities are considered, R&D is in many cases not clear. In addition, the issue of "change of intent" is debated, as it is often hard to clearly differentiate between noncommercial and commercial use. Even when jurisdictions provide guidance and interpretation in this regard (*6*), uncertainty will remain for curators of international biobanks because interpretation of the NP's scope can differ elsewhere in the world. This leaves biobanks in the dark about the legal status and terms of use of their collections.

Another problem is that implementation of the NP conditions in individual countries' national legal systems is a time-consuming process that has not yet been completed in most countries that ratified the NP, 4 years after it entered into force (*7*). Inclusion of GSD into the NP could lead to even more uncertainty. Opinions on this seem to be deeply divided, starting with widespread confusion about the definition of "digital sequence information" that is now broadly used by the NP (*8*).

The combination of partial implementation of the NP, lack of clarity on rules and processes, and differences in interpretation at a national level are foreseeable major hurdles, especially in outbreak situations when the timeliness of access to PGR is of essence. Unless clarity is provided for such matters, the ability to respond to outbreaks that require international data sharing is likely to be severely hampered. These barriers are similar for individual researchers and biobanks, but the practical impact for large collections is more substantial, as due diligence has to be exercised with each sample—and possibly soon, each GSD—by acquiring bilateral agreements and necessary documentation determining what is permitted on the grounds of the specific MAT.

FOUR MODELS FROM BIOBANKS

We present here four models, differing in the role and responsibility of the biobank in negotiating and processing the NP conditions concerning PIC and MAT. We searched examples of biobanks and biobank networks that admit in their collections multiple pathogens and supply materials to multiple stakeholders in different countries. The main difference between the models is where the burden to

negotiate ABS conditions lies: Some biobanks and networks require this responsibility from each person (and institute) depositing the sample; some take up this responsibility themselves, providing services and harmonizing procedures; and some simply provide the offered materials while leaving the responsibility to acquire PIC and MAT to (end)users (see the table and figs. S1 and S2).

The American ATCC and Japanese NITE are models of nationally curated biobanks (see fig. S1). In the ATCC, no provision for the acquisition and transfer of NP conditions is made, because the United States has not signed on to the CBD. Thus, users still have to acquire bilaterally all NP documentation themselves (*7*, *9*). The time-consuming and costly bilateral negotiations with multiple countries at a governmental level is problematic for the rapid availability of PGR necessary in a public health emergency. By contrast, the NITE offers certainty about the NP conditions of the materials through preestablishment of bilateral PIC and MAT agreements with certain Asian countries to facilitate transfers of genetic resources. To our knowledge, this is the only model in which a biobank takes up the NP burden also for commercial use of materials [though only for Japanese stakeholders (*10*)].

The Asian ACM and European ECCO models consist of a network of multiple national biobanks (see fig. S2). In the ACM, there is no need to negotiate bilaterally PIC and MAT, because each of the 13 governments from the countries in the network entered into a memorandum of understanding (MoU) with the ACM curator. A MoU is, in short, not a strictly legally but rather a politically binding agreement. In this case, individual countries' PIC and MAT conditions are waived and replaced by a simple, standardized set of ABS conditions. The NP burden is in this way facilitated by the network curator and the result is, in principle, rapid sharing of materials for noncommercial use between network partners (*11*). By contrast, for the ECCO model, the administrative NP burden stays on each depositing party that has to submit the PIC and MAT certificates together with the materials at the beginning. The (network of) biobanks only control the existence and transfer of the NP certificates. When transferring to other repositories or end users, a model MTA of ECCO must be used, but each time the different MAT of each sample has to be incorporated (*12*, *13*). Two challenges still remain for both network models: (i) defining the line between noncommercial and commercial activities and (ii) how public health activities are interpreted in this regard.

REDUCE BARRIERS FOR PUBLIC HEALTH

The American and European models require, either at the start (ECCO) or at the end (ATCC) of the process, isolate-specific bilateral negotiations with the government of origin, resulting in (i) a time-consuming process and (ii) a collection in which each isolate has different conditions of use. The Japanese NITE and Asian ACM models, although containing facilitated ABS negotiations and standard conditions for noncommercial use of materials, consist of closed networks in which only members benefit from the facilitated channels. The standard MTA for noncommercial use in countries outside the ACM network is, however, notable: After signing an MoU, individual PIC and MAT conditions are waived to allow the use of an agreed standard MTA. This is a legally binding contract, resulting in a multilateral system instead of many bilateral negotiations.

In addition to the above-described models, we highlight two multilateral examples for negotiating ABS conditions, developed to work on a global scale: the Food and Agriculture Organization of the United Nations (FAO) International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) developed in 2001, before the adoption of the NP, and later recognized under Article 4 of the NP (*5*); and the WHO Pandemic Influenza Preparedness (PIP) Framework, adopted in 2011.

The ITPGRFA uses a standard MTA that covers commercial and noncommercial use. In this MTA, tailor-made benefit-sharing obligations are imposed on the individual users, as indicated by each providing country, beforehand, in a contract with the centralized administrator or curator (*14*). Further research is needed to assess whether this model could work efficiently and quickly for pathogens, and at what cost. The PIP Framework, by contrast, applies two different standard MTAs for the different intents (commercial versus noncommercial) of use of influenza samples. An in-kind (vaccine supply) and a monetary donation are agreed with "industry partners" (vaccine, diagnostic, and pharmaceutical manufacturers) to be collected by the WHO from the commercial users (*15*).

The PIP Framework has been successful but has endured several challenges. Laboratories outside the WHO Global Influenza Surveillance and Response System are considered potential commercial users by default, even if they have a public health mandate. The framework also requires a costly administrative and stewardship system of $6 million per year (*15*). In light of such costs, and questions about whether other pathogens have the potential commercial value of influenza, developing similar pathogen-specific

National and regional biobank models for compliance with the Nagoya Protocol conditions

COUNTRIES AND NETWORKS	TYPE OF FUNDING	ACQUISITION REQUIREMENTS	COMMERCIAL TRANSFERS	NON-COMMERCIAL TRANSFERS
American Type Culture Collections (ATCC); United States of America	Fee-for-service	Agreements on ownership transfers	License for commercial use and bilateral negotiations with NP countries	MTA from ATCC and bilateral negotiations with NP countries
Japanese National Culture Collection [National Institute of Technology and Evaluation (NITE)]; bilateral agreements with Indonesia, Myanmar, Vietnam, Mongolia, Thailand, China, and Korea	Publicly funded (Japanese government)	Project agreements (based on MoU), PIC and MAT	Bilateral MTAs, pre-negotiated by NITE for Japanese users	Standard MTA ACM-NIEMA (Network of International Exchange of Microbes under the ACM)
The Asian Consortium for the Conservation and Sustainable Use of Microbial Resources (ACM); Cambodia, China, India, Indonesia, Japan, Korea, Laos, Malaysia, Mongolia, Myanmar, Philippines, Thailand, and Vietnam	Publicly funded (member countries)	Memorandum of Understanding where PIC/MAT are waived	Not covered by the system. Users have to obtain PIC and MAT, and negotiate bilateral MTAs with the Nagoya countries	Material transfer form for members, and standard MTA ACM-NIEMA for nonmembers
European Culture Collections' Organization (ECCO); Austria, Belgium, Bulgaria, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Italy, Latvia, Norway, Poland, Portugal, Romania, Russia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Netherlands, Turkey, and United Kingdom	Publicly funded (member countries)	PIC and MAT	Not covered by the system. Users have to obtain PIC and MAT, and negotiate bilateral MTAs with the Nagoya countries	Model MTA ECCO with a conditional clause that allows redistribution only for noncommercial use

multilateral frameworks for pathogens other than influenza does not seem like a sustainable option (*7*).

Another way of approaching the undesired delay of ABS negotiations, especially for access and use of PGR during public health emergencies, is the creation of exemptions for this specific context. There is no clear guidance in the NP for the fast sharing of genetic resources in emergency situations, besides a general call on countries to pay due regard to health-threatening situations, in which they may consider in their domestic regulation the need for expedited ABS provisions [Article 8 of the NP (*5*)]. The European Union (EU) regulation on compliance with the NP (*6*) contains a special provision for the sharing of PGR in public health emergencies, including the role of biobanks, as in the ECCO model, as "trusted intermediaries." The key feature is the exemption from the NP's due diligence conditions for a period of 3 months, for strains that are likely to be the cause of a (possible) public health emergency of international concern. After these 3 months, PIC and MAT have to be negotiated or all the materials that were shared destroyed.

In support of this provision, EU guidance documents mention the need for a fast-track procedure, in which access to PGR may be granted first and PIC and MAT acquired later. As such, formal consent can be given retrospectively based on a tracking system with an attached electronic tag known as the global unique identifier. How this exemption period will actually work and whether it will be good enough remains to be seen. Outbreaks can last for a long time, and on ethical grounds, the development of state-of-the-art diagnostics or pharmaceutical products is not easily stopped or turned back when people are seriously at risk or suffering from diseases. In any case, considering an exemption of PGR to the NP should by no means imply that benefits are withheld; such benefits must be shared at least in the form of nonmonetary collaborations and support to epidemic control and response. Outbreak-related research and countermeasures, such as diagnostics and vaccines, must be accessible to all affected countries not only as a legal obligation, but also as an ethical imperative.

If GSD will come under the scope of the NP, a substantial increase of complex administrative mechanisms will arise, both for the receiving party and the providing countries, as well as intermediaries such as biobanks and databases. Additional to managing the huge volume of data, a diversity of hard-to-resolve dilemmas about definitions and the applicability of PIC and MAT to GSD can be foreseen. These include disputes about the origin of the data due to difficulties in tracking transfers; the legality of use of partial sequences; confusion on the status of results of metagenomics and reverse genetics; ownership issues on incremental innovation; and compliance issues for open science, funders, and editors, requiring GSD to be fully and publicly available (*8*).

CLARIFICATION, COLLABORATION

With the increasing threat of globally dispersed outbreaks, a global response is of crucial importance; hence, the scattered interpretations and implementations of the NP can seriously affect global health by hampering the timely sharing of essential PGR (*8*).

We found no single biobank or network model for compliance with the NP that resolves all challenges for biobanks to facilitate access, rapid transfer, and use of genetic resources in all situations and on a global scale. The inclusion of GSD in such models will only increase complexity and legal uncertainty, adding to the already persistent problem of tracking the access and use of materials and isolates. Monitoring and traceability of access, transfer, and use of the enormous, and increasing, volume of data are practically unfeasible (*8*). Additionally, as the nature and extent of use of GSD are variable, defining benefits from the use of the data can be extremely challenging. The NP was not developed specifically by and for the biomedical field, but for broad biodiversity protection and for fair development, trade, and intellectual property objectives, which are usually addressed in separate disciplines and debated in distinct political institutions. This influences the ability of the NP to address comprehensively biomedical needs and practices, especially during public health emergencies.

"...the scattered interpretations and implementations of the Nagoya Protocol can seriously affect global health..."

For swift outbreak response, the world needs to tackle the challenges of excessive ownership over genetic resources, possibly by transferring due diligence obligations to later stages of product development, alleviating the burden and consequential delay on their use for basic research and public health surveillance and response. Further needs are the implementation of simplified sharing agreements based on standardized, multilateral ABS procedures, harmonizing disparate databases and platforms, and clearly defining which activities fall under the NP's scope, including how "digital sequence information" will be understood and treated, with possible exemptions, at least in cases of public health emergencies. Only then will research laboratories and public health institutes be able to contribute optimally to global collaborations for state-of-the-art diagnostics and swift surveillance and response in public health emergencies. ■

REFERENCES AND NOTES

1. S. A. Laird, R. P. Wynberg, The emergence and growth of digital sequence information in research and development: Implications for the conservation and sustainable use of biodiversity, and fair and equitable benefit sharing. A fact-finding and scoping study undertaken for the Secretariat of the Convention on Biological Diversity (2017).
2. United Nations, UN Convention on Biological Diversity (1992); www.cbd.int/convention/text/default.shtml.
3. Convention on Biological Diversity, Fourteenth meeting provisional agenda (2018); www.cbd.int/doc/c/0d16/208c/39321b05c9f6aba8512d597f/cop-14-01-en.pdf.
4. K. Modjarrad *et al.*, *PLOS Med.* **13**, e1001935 (2016).
5. Secretariat of the Convention on Biological Diversity, The Nagoya Protocol on access to genetic resources and the fair and equitable sharing of benefits arising from their utilization to the convention on biological diversity, Montreal (2011).
6. Regulation (EU) No 511/2014 of the European Parliament and of the Council of 16 April 2014 on compliance measures for users from the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization in the Union (2014).
7. J. H. Reichman, P. F. Uhlir, T. Dedeurwaerdere, *Governing Digitally Integrated Genetic Resources, Data, and Literature: Global Intellectual Property Strategies for a Redesigned Microbial Research Commons* (Cambridge Univ. Press, 2016).
8. Convention on Biological Diversity, AHTEG on Digital Sequence Information on Genetic Resources. Peer review of the fact-finding and scoping study and Submissions from Parties, other Governments, relevant organizations and stakeholders; www.cbd.int/abs/dsi-gr/ahteg.shtml.
9. ATCC. Material Transfer Agreement (American Type Culture Collection, 2011); www.lgcstandardsatcc.org/Documents/Product_Use_Policy/Material_Transfer_Agreement.
10. NITE, Bilateral Relation with Asian Nations (National Institute of Technology and Evaluation, 2017); www.nite.go.jp/en/nbrc/global/asia/about_egron.html.
11. K. Ando *et al.*, *Microbiol. Cult. Collect.* **30**, 85 (2014).
12. ECCO core Material Transfer Agreement for the supply of samples of biological material from public collections (2009).
13. TRUST. Transparent Users-Friendly System of Transfer (2016).
14. J. H. Reichman, Why the Nagoya Protocol to the CBD Matters to Science and Industry in Canada and the United States (2018).
15. WHO, External evaluation of the Pandemic Influenza Preparedness Partnership Contribution – High-Level Implementation Plan 2013-2016; www.who.int/about/evaluation/pip_evaluation_report.pdf?ua=1.

ACKNOWLEDGMENTS

We acknowledge the members of the COMPARE Consortium and the European partner project European Virus Archive goes global (EVAg) for inspiration and support. **Funding:** This study was done under the framework of the EU COMPARE project. This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no. 643476. **Author contributions:** All authors participated in developing the article's concept and in the critical revision of the manuscript. C.d.S.R. and G.B.H. conceptualized and produced the table and figures. C.d.S.R. drafted the article text, which was revised and edited by G.B.H. and M.P.K. All authors approved the final version of the manuscript and agreed to be accountable for all aspects of the work. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6413/404/suppl/DC1

10.1126/science.aau5229

BOOKS *et al.*

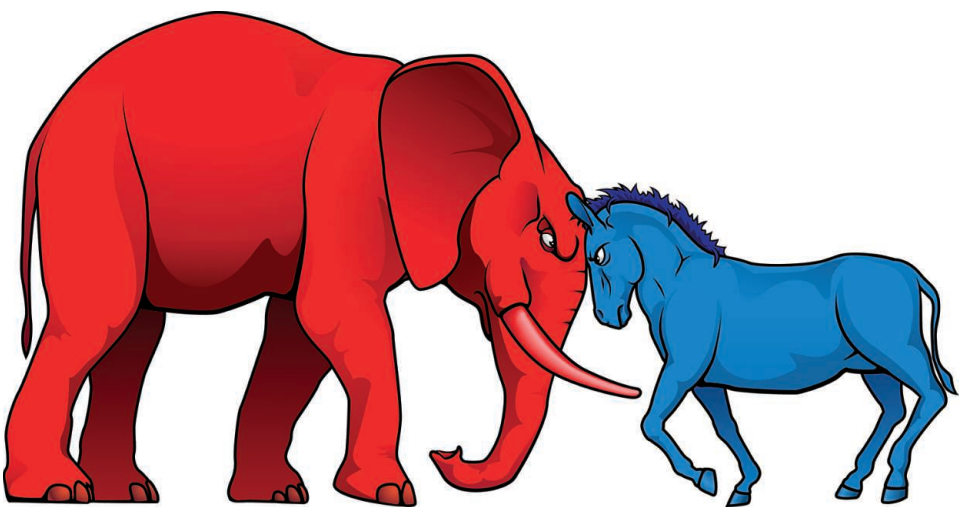
PARTISANSHIP

Us against them

Is contemporary partisanship "identity all the way down"?

By **Christopher M. Federico**

If there is a single buzzword that sums up American politics in the present era, it is "polarization." Though citizens and commentators alike loosely use the term to refer to the acrimony that characterizes relations between Democrats and Republicans, political scientists often have something very specific in mind when they talk about polarization. These days, that "something" is the tendency for the two political parties to strongly dislike one another—what scholars refer to as "affective polarization."



But why has affective polarization become such a prominent feature of American political life? In *Uncivil Agreement*, Lilliana Mason provides a novel and compelling answer.

By way of background, a common explanation for why Democrats and Republicans have come to dislike one another so much is that they take starkly different views on the issues. This would seem to make sense: If your political goals differ from mine, I am less likely to feel warmly about you.

Mason takes issue with this explanation and instead suggests that polarization is rooted in a primal sense of team identification. For many, she argues, being a Democrat or a Republican has become increasingly important to their sense of self. If I identify psychologically with a particular party, my animosity toward the other political "team" and desire for a partisan "victory" become ends in themselves—even when I don't have especially strong feelings about a specific issue on which we disagree.

Mason argues that party identities have become powerful because they overlap more and more with other important identities, such as race, religious affiliation, or thinking of oneself as a liberal or a conservative. As she puts it, partisans have become "socially sorted": Liberals, people of color, and the less religious increasingly identify as Democrats, whereas conservatives, whites, and the more religious increasingly identify as Republicans. Thus, partisan affiliations now also imply differences in ideology, race, and religion, reinforcing our sense of distance from, and competition with, those who belong to the opposing party.

Mason fleshes this argument out over the course of an eminently readable narrative. After laying out the book's basic premise in the first few chapters, she provides descriptive evidence that Americans have socially sorted themselves into different parties as a function of ideology, race, and religion and that they have become more hostile to those on the other partisan team. She then documents the consequences of these developments.

In Mason's fifth chapter, perhaps the most important in the book, she provides evidence that it is socially sorted partisans rather than partisans with extreme issue positions who are more likely to dislike members of the other party and show biases in favor of their own party. For example, Democrats who also label themselves as liberals are more likely to dislike Republicans, even when those same Democrats express relatively moderate or even conservative issue positions. In subsequent chapters, Mason expands on this point, demonstrating that citizens whose partisan identities overlap with their ideology, race, or religious orientation are more likely to express anger toward members of the other party and are more likely to bring that anger into the public arena by participating in politics more avidly.

She concludes by reviewing the implications of her argument for the future of American democracy. Though Mason is circumspect about how easily we might tamp down the identity-fueled conflict currently at the heart of American politics, she does discuss several scientifically informed steps for us to consider. Some of these are classic remedies for intergroup conflict emphasized by psychologists, such as increased contact between Democrats and Republicans and finding shared goals or identities that can unite people across partisan lines. Others focus on possible changes in the parties themselves, including a stronger emphasis among party leaders in setting norms of comity and tolerance and the prospect of intensified divisions in the Republican Party that introduce "cross-cutting cleavages that suppress social polarization and social distance."

Mason's book is both enjoyable to read and invaluable for making sense of our polarized politics. Above all, it hammers home the overriding importance of social identity for how citizens navigate the political world. Though this has been understood by psychologists for some time (*1*), it has only recently begun to be fully appreciated by students of political behavior. ■

The reviewer is at the Departments of Political Science and Psychology, University of Minnesota, Minneapolis, MN 55455, USA. Email: federico@umn.edu

Uncivil Agreement
How Politics Became Our Identity
Lilliana Mason
University of Chicago Press, 2018. 192 pp.

REFERENCE

1. H. Tajfel, J. C. Turner, in *Psychology of Intergroup Relations*, S. Worchel, W. G. Austin, Eds. (Nelson-Hall, 1986), pp. 7–24.

10.1126/science.aav1606

POLITICAL ATTITUDES

The political legacy of slavery

Racial tension continues to influence political attitudes in U.S. regions that once depended on slave labor

By **Amber Spry**

Debates about immigration, the emergence of #BlackLivesMatter and NFL activism, and the lingering presence of alt-right and white supremacy movements signal that racial attitudes still color interactions in the United States. Amid this backdrop, *Deep Roots* emerges as a story about the determinants of white racial attitudes in the South, a region known for being more ideologically and politically conservative than the rest of the country, particularly on racial issues.

The book argues that political culture in the South can be explained by the path dependence of political attitudes—that is, that ideas, norms, and behaviors of the past can influence the present. The argument sits in dialogue with a long-standing literature in institutional path dependence but differs in its approach by making political attitudes and behaviors, rather than institutional outcomes, the subject of observation. The authors make an important contribution by showing that just as attitudes and behavioral norms can be passed down through families and social circles, so too can they be reinforced by institutions.

Authors Avidit Acharya, Matthew Blackwell, and Maya Sen begin *Deep Roots* by highlighting two southern cities: Greenwood, Mississippi, and Ashville, North Carolina. Situated in the Mississippi Delta, Greenwood has a little more than 15,000 residents, the majority of whom are African American and vote predominantly for the Democratic Party. Greenwood's white residents, who are in the minority, are overwhelmingly conservative. In contrast, Ashville, a city with around 83,000 residents, has a predominantly white population that is more liberal by comparison.

The authors use these two cities to reveal what they believe to be a broader pattern in the present-day politics of the American South: Some areas have grown relatively liberal in their politics while other areas—particularly areas along the "Black Belt" region, where slavery was most prevalent—remain among the most ideologically and politically conservative places in the country.

These differences, the authors argue, exist because cities like Greenwood were places where the local economy was rooted in slavery, whereas in places like Ashville it was not. Through survey analyses, they show that in areas of the South where slavery was prevalent, white respondents were less likely to identify as a Democrat, less supportive of affirmative action, and more likely to express racially resentful attitudes.

The book provides a detailed account of how behavioral paths in the South diverged, supported by quantitative analysis and compelling anecdotal evidence, and explains why these trajectories were not disrupted by events such as the passage of the Civil Rights Act of 1964 and the Voting Rights Act of 1965. And although the authors show that outcomes did improve for minorities in the Black Belt after both acts were passed (more African Americans were registered to vote, educational attainment increased, and gaps in income inequality narrowed), the political attitudes and behaviors of whites in this region remained conservative.

The authors reassure readers that targeted policy interventions can attenuate the effects of slavery on important outcomes related to social, political, and economic inequality between white and black Americans. But to the related question of whether these policy interventions have impacted latent attitudes among conservative white Americans, the book's evidence suggests that negative racial attitudes held by southern whites about black Americans are remarkably persistent.

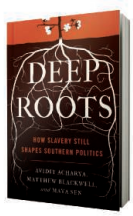
Although the analysis in *Deep Roots* is limited in scope to the American South, the theory of behavioral path dependence presents several ways forward for continued work to examine the contours of political attitudes, both in the United States and in comparative contexts. For example, I wonder whether behavioral path dependence is a useful framework for understanding race-related political attitudes on issues such as immigration. Additionally, I am curious to know the extent to which behavioral path dependence also helps explain more positive attitudinal developments, or whether the emergence and persistence of progressivism requires a different theory.

Scholars of racial attitudes have long considered how such attitudes are transmitted across generations through history, culture, and institutions, and *Deep Roots* makes a historically penetrating and theoretically meaningful contribution to that body of literature. The book is engaging and thorough in its analysis and puts forth theory that will be useful for readers specifically interested in the intersections of political geography, racial attitudes, and political behavior. ■

10.1126/science.aav1297

The reviewer is at the Departments of Politics and African and African American Studies, Brandeis University, Waltham, MA 02453, USA. Email: amberspry@brandeis.edu

Deep Roots
How Slavery Still Shapes Southern Politics
Avidit Acharya, Matthew Blackwell, Maya Sen
Princeton University Press, 2018. 296 pp.



Barack Obama hosts an online virtual town hall in 2009.

PHOTO: BROOKS KRAFT LLC/CORBIS/GETTY IMAGES

PODCAST

Politics with the People
Building a Directly Representative Democracy
Michael A. Neblo, Kevin M. Esterling, David M. J. Lazer
Cambridge University Press, 2018. 182 pp.

Town hall meetings in the United States tend to attract politicians' biggest supporters and their most vocal critics, leaving out a wide swath of constituents. But with the advent of online communication platforms, that needn't remain the case. This week on the *Science* podcast, David Lazer outlines a technology-enabled framework for creating a more representative democracy.

10.1126/science.aav5181

Calls to protect the Mauritian flying fox (*Pteropus niger*) have gone unheeded.

PHOTO: JACQUES DE SPEVILLE

LETTERS

Edited by **Jennifer Sills**

Broader conservation strategies needed

Over the past 3 years, the Mauritian flying fox (*Pteropus niger*), a Mascarene endemic and threatened island fruit bat, has been culled to half its global population in attempts to increase fruit producers' profits (*1*). As a result, its Red List status was recently reassessed from Vulnerable to Endangered (*2*). Culling of flying foxes has been shown to be ineffective in boosting profits of commercial fruit producers; lychee production dropped by some 70% (*3*) after the previous mass culling campaigns. Despite the risks, the lack of benefits, and the existence of effective alternatives (*4*), Mauritius is planning a new mass cull for 2018 (*5*).

The fruit bat is the last survivor of three *Pteropus* species and of other large frugivores such as Dodos and giant tortoises that used to inhabit the island (*6*). By eating fruits and potentially disseminating seeds of about half of the native trees of the island's forests (*7*), the species fulfills a keystone ecological role in Mauritius's remnant native habitats. Only 5% of these habitats remain, and they harbor one of the most threatened biota worldwide (*8*).

Mauritius's biodiversity protection law was weakened in 2015 specifically to enable mass culls (*9*), which are supported by National Parks and Conservation Service resources that were originally intended to conserve the country's threatened biodiversity (*10*). Events unfolding in Mauritius are exposing the limitations to the approach local and international conservationists and conservation organizations have used as they call for reasonable management. Evidence, appeals, petitions, articles (*11*), street protests, and the missions of international experts such as the International Union for Conservation of Nature (*10*) have been largely ignored.

Conservationists should diversify and intensify their approaches for tangible results by incorporating litigation that stalls unnecessary biodiversity destruction (*12*). For example, the 2015 mass cull breached the law in place at the time and therefore may have been prevented by an injunction (*9*). In the longer term, conservationists should apply marketing persuasion principles to conservation. They should focus on educating the public, fruit producers, and ministry advisers and technicians about the ineffectiveness of culling and the nonlethal alternative solutions. Conservationists should also identify and address barriers to the implementation of evidence-based policy, including the focus on short-term goals driven by election cycles at the expense of long-term environmental interests (*12*).

F. B. Vincent Florens[1]* and Christian E. Vincenot[2]
[1]University of Mauritius, Réduit, Mauritius.
[2]Department of Social Informatics, Kyoto University, Kyoto, Japan.
*Corresponding author.
Email: vin.florens@uom.ac.mu

REFERENCES

1. C. E. Vincenot *et al.*, *Science* **355**, 1368 (2017).
2. T. Kingston, F. B. V. Florens, R. Oleksy, K. Ruhomaun, V. Tatayah, *Pteropus niger* (The IUCN Red List of Threatened Species, 2018); www.iucnredlist.org/details/18743/0.
3. "Letchis : La récolte baisse de 70%," *Business Mega* (2017); https://business.mega.mu/2017/11/03/letchis-la-recolte-baisse-de-70/ [in French].
4. S. A. Aziz *et al.*, in *Bats in the Anthropocene: Conservation of Bats in a Changing World*, C. C. Voigt, T. Kingston, Eds. (Springer International AG, 2016), pp. 377–426.
5. "Nouvelle campagne d'abattage de la chauve-souris à l'agenda," *IONNEWS* (2018); http://ionnews.mu/nouvelle-campagne-dabattage-de-la-chauve-souris-mauricienne-a-lagenda-17072018/ [in French].
6. D. M. Hansen, M. Galetti, *Science* **324**, 42 (2009).
7. F. B. V. Florens *et al.*, *J. Nat. Conserv.* **40**, 85 (2017).
8. A. Cheke, J. Hume, *Lost Land of the Dodo: An Ecological History of the Mascarene Islands* (T and AD Poyser, London, 2008).
9. F. B. V. Florens, *Nature* **530**, 33 (2016).
10. Republic of Mauritius, Sixth National Assembly, Parliamentary Debates (2016), pp. 50-51; http://mauritiusassembly.govmu.org/English/hansard/Documents/2016/hansard0816.pdf.
11. J. Aldred, "Conservationists urge Mauritius to halt cull of threatened fruit bat," *The Guardian* (2015); www.theguardian.com/environment/2015/nov/17/conservationists-urge-mauritius-to-halt-cull-of-threatened-fruit-bat.
12. G. Chapron, Y. Epstein, A. Trouwborst, J. V. López-Bao, *Nat. Ecol. Evol.* **1**, 0086 (2017).

10.1126/science.aav5161

Protect NIH's DNA advisory committee

The Recombinant DNA Advisory Committee (RAC) was founded at the behest of concerned scientists based on the principle that public trust of advanced technologies must be earned through transparency and expert deliberation. For more than 40 years, the RAC has served as custodian of the National Institutes of Health (NIH) guidelines, with responsibilities in human gene therapy evolving from regulatory to advisory to the NIH and Food and Drug Administration (FDA). However, the NIH director and FDA commissioner have proposed the elimination of RAC review of human gene therapy protocols (*1*, *2*).

The changes proposed to the NIH guidelines are even broader: "Eliminat[e] references to the RAC, including its roles in [human gene therapy] and biosafety" (*3*). The NIH and FDA cited a lack of recent RAC reviews (three in 2 years) as justification for the changes (*1*). Yet, since the NIH accepted the Institute of Medicine's recommendations to restrict the RAC's advisory role (*4*), NIH itself, not the RAC, is responsible for setting the RAC's agenda (*5*). Removing RAC involvement from all aspects of human gene therapy and laboratory biosafety continues a worrisome trend of cutting independent experts and public discussion out of government decision-making (*6*).

ClinicalTrials.gov has been proposed as an alternative to the RAC (*1*). However, this system does not extend to laboratory research, provides insufficient information about ongoing human gene therapy trials to satisfy the public's interest in transparency, and—unlike the RAC—involves no scientific oversight (*7*). Current regulations require only that clinical trial results be posted within 1 year of trial completion (*8*), and in practice, this information is rarely posted.

Local institutional biosafety committees and institutional review boards, whose expertise varies widely, now assume sole responsibility for assessing the hazards and ethics of research using emerging recombinant DNA technologies and planning appropriate steps for their mitigation. At a recent meeting convened by the NIH Office of Science Policy, representatives of institutional biosafety committees voiced concern over these responsibilities and asked instead for better guidance on new emerging technologies from the NIH and the RAC (*9*, *10*). We urge the NIH director to reconsider the consequences of the proposed changes and ensure that viable alternatives are in place that preserve the transparency and accountability required to maintain public trust in the regulation of recombinant DNA, laboratory biosafety, and human gene therapy (*11*).

Zach N. Adelman,[1]*† Lorraine M. Albritton,[2]† Kathleen Boris-Lawrie,[3]† Michael J. Buchmeier,[4]‡ Paula Cannon,[5]‡ Mildred Cho,[6]† David DiGiusto,[7]† J. Kevin Donahue,[8]‡ Howard J. Federoff,[9]‡ Marie-Louise Hammarskjold,[10]‡ Angelica D. Hardison,[11]‡ Patrick Hearing,[12]‡ Benhur Lee,[13]† Dean Anthony Lee,[14]† Matthew H. Porteus,[15]† Lainie Friedman Ross,[16]‡ Susan R. Ross,[17]‡ Dawn P. Wooley,[18]‡ Laurie Zoloth[19]‡

[1]Texas A&M University, College Station, TX 77843, USA. [2]College of Medicine, The University of Tennessee Health Science Center, Memphis, TN 38163, USA. [3]University of Minnesota, Twin Cities, MN 55108, USA. [4]Division of Infectious Disease, University of California, Irvine, Irvine, CA 92697, USA. [5]Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA. [6]Stanford University, Stanford, CA 94305-5417, USA. [7]Semma Therapeutics, Cambridge, MA 02142, USA. [8]University of Massachusetts Medical School, Worcester, MA 01655, USA. [9]College of Medicine, University of California, Irvine, Irvine, CA 92617, USA. [10]Myles H. Thaler Center for AIDS and Human Retrovirus Research, University of Virginia, Charlottesville, VA 22908, USA. [11]Health System Office for Compliance and Enterprise Risk Management, Augusta University, Augusta, GA 30912, USA. [12]School of Medicine, Stony Brook University, Stony Brook, NY 11794-5222, USA. [13]Icahn School of Medicine at Mount Sinai, New York, NY 10029-6508, USA. [14]Nationwide Children's Hospital, Columbus, OH 43110, USA. [15]Center for Definitive and Curative Medicine, Stanford University, Stanford, CA 94305, USA. [16]MacLean Center for Clinical Medical Ethics, University of Chicago, Chicago, IL 60637, USA. [17]College of Medicine, University of Illinois at Chicago, Chicago, IL 60612, USA. [18]Wright State University, Dayton, OH 45435, USA. [19]Divinity School, University of Chicago, Chicago, IL 60637, USA.
*Corresponding author. Email: zachadel@tamu.edu
†RAC member
‡Past RAC member

REFERENCES AND NOTES

1. F. S. Collins, S. Gottlieb, *N. Engl. J. Med.* **379**, 1393 (2018).
2. J. Brainard, *Science* **361**, 734 (2018).
3. NIH, "National Institutes of Health (NIH) Office of Science Policy (OSP) recombinant or synthetic nucleic acid research: Proposed changes to the NIH guidelines for research involving recombinant or synthetic nucleic acid molecules (NIH Guidelines)" (2018); www.federalregister.gov/d/2018-17760/p-14.
4. R. N. Lenzi, B. M. Altevogt, L. O. Gostin, Eds., *Oversight and Review of Clinical Gene Transfer Protocols* (National Academies Press, 2018).
5. NIH, "Final action under the NIH Guidelines for Research involving recombinant or synthetic nucleic acid molecules (NIH Guidelines)" (2016); www.federalregister.gov/d/2016-06448.
6. K. Boyle, M. Kotchen, *Science* **361**, 729 (2018).
7. NIH, About the Results Database (https://clinicaltrials.gov/ct2/about-site/results).
8. NIH, "2016: Final NIH policy on the dissemination of NIH-funded clinical trial information issued" (https://clinicaltrials.gov/ct2/about-site/history#FinalNIHPolicy).
9. NIH, "NIH guidelines: Honoring the past, charting the future" (2017); https://osp.od.nih.gov/event/nih-guidelines-honoring-the-past-charting-the-future.
10. NIH, "NIH guidelines: Honoring the past, charting the future workshop (day 2)" (2017); https://videocast.nih.gov/summary.asp?live=24766&bhcp=1.
11. These views represent the authors' own opinions as members of the scientific community, not the views of the RAC.

10.1126/science.aav2483

A proposed U.S. law would allow researchers to make better use of agricultural data.

PHOTO: SCOTT OLSON/GETTY IMAGES

The power of agricultural data

Federal agricultural data collection and management in the United States is decentralized and fragmented, hampering the ability of policy-makers to empower growers and researchers in data analytics. A bill introduced by Senators John Thune (R–SD) and Amy Klobuchar (D–MN) and included in the Senate farm bill could help. Bipartisan bill S. 2487, the Agriculture Data Act of 2018 (*1*), directs the U.S. Department of Agriculture (USDA) to establish a secure data warehouse and a protocol for confidential data sharing with university researchers to allow for better use of the USDA's vast data resources. This would unlock enormous opportunities for data-driven scientific discovery that will eventually lead to better conservation outcomes.

Broadly linking historic microdata across the USDA will lead to research and tools that could better enable producers and policy-makers to identify and operationalize activities that improve water quality, risk management policies, nutrient effluence management, fertility and pesticide use, and in-field management (*2*–*6*). Pollution mitigation and economic growth are not mutually exclusive (*7*), but better data analytics and use of existing data are necessary to bridge the gap. Rapid advancements in georeferenced technologies and analytics offer an unprecedented opportunity to use these data to enable smarter production systems.

The Agriculture Data Act of 2018 would be an important first step to harness the power of USDA data, providing researchers the ability to conduct big-data–oriented research in agriculture, the environment, nutrition, and supply chains. Such research would enable the design of policies that inspire changes in agricultural practices, improve producer profitability, and lead to better environmental outcomes. Strongly held concerns about data privacy have inspired frequent conversations among producers,

researchers, conservation groups, congressional leaders, and USDA legal counsel experts about best approaches to data integration. The proposed Agriculture Data Act of 2018 ensures the protection of private data while unleashing the power of data that are currently siloed by the federal government. We encourage its passage.

Joshua D. Woodard,[1]* Bruce J. Sherrick,[2] Deborah M. Atwood,[3] Robert Blair,[4] Greg Fogel,[5] Nicholas Goeser,[6] Barry Gold,[7] Josette Lewis,[8] Carl Mattson,[9] Jim Moseley,[3] Collin O'Mara,[10] John Piotti,[11] Bill Salas,[12] Lynn Scarlett,[13] Kristin Weeks Duncanson,[14] Fred Yoder[15]

[1]Dyson School of Applied Economics and Management, Cornell University, Ithaca, NY 14853, USA. [2]College of Agricultural, Consumer, Environmental Sciences, University of Illinois at Urbana, Urbana, IL 61801, USA. [3]AGree | Transforming Food and Ag Policy, Washington, DC, 10026, USA. [4]Three Canyon Farms, Kendrick, ID 83537, USA. [5]National Sustainable Agriculture Coalition, Washington, DC 20002, USA. [6]National Corn Growers Association, Chesterfield, MO 63005, USA. [7]Walton Family Foundation, Washington, DC 20006, USA. [8]Environmental Defense Fund, Washington, DC 20009, USA. [9]George Mattson Farms Inc., Chester, MT 59522, USA. [10]National Wildlife Federation, Reston, VA 20190, USA. [11]American Farmland Trust, Washington, DC 20036, USA. [12]Applied GeoSolutions LLC, Durham, NH 03824USA. [13]The Nature Conservancy, Arlington, VA 22203, USA. [14]Highland Family Farms, Mapleton, MN 56065, USA. [15]The North American Climate Smart Agriculture Alliance, Lutherville, MD 21093, USA. *Corresponding author. Email: josh@ag-analytics.org

REFERENCES

1. Agriculture Data Act of 2018, S. 2487, 115th Cong. (2017–2018).
2. J. D. Woodard, *Environ. Sci. Pol.* **66**, 93 (2016).
3. J. D. Woodard, *Appl. Econ. Perspect. Pol.* **38**, 373 (2016).
4. K. A. Garrett, *Nat. Clim. Change* **3**, 955 (2013).
5. D. Gustafson *et al.*, *BioScience* **66**, 80 (2016).
6. B. Phalan *et al.*, *Science* **351**, 450 (2016).
7. B. Obama, *Science* **355**, 126 (2017).

Competing interests: J.D.W., J.M., N.G., B.S., C.M., K.W.D., R.B., F.Y., and D.M.A. are members of AGree's Conservation and Crop Insurance Taskforce. J.D.W. provided unpaid technical advice to U.S. Senate staff responsible for drafting the Agriculture Data Act of 2018 legislation, which was informed by AGree's Conservation and Crop Insurance Task Force. J.D.W. operates Ag-Analytics.Org, a farm management and data platform, and is an expert reviewer for the Federal Crop Insurance Program as nominated by the chief economist of the USDA and authorized by the Federal Crop Insurance Corporation board of directors.

10.1126/science.aav5002

TECHNICAL COMMENT ABSTRACTS

Comment on "The earliest modern humans outside Africa"

Warren D. Sharp and James B. Paces

Hershkovitz *et al*. (Reports, 26 January 2018, p. 456) interpreted the Misliya-1 fossil maxilla as evidence of the earliest known anatomically modern human outside Africa. However, the fossil's reported age of 177,000 to 194,000 years relies on flawed interpretations of uranium-series data. We contend that those data support a minimum age of no more than ~60,000 to 70,000 years.

Full text: dx.doi.org/10.1126/science.aat6598

Response to Comment on "The earliest modern humans outside Africa"

Israel Hershkovitz, Mathieu Duval, Rainer Grün, Norbert Mercier, Helene Valladas, Avner Ayalon, Miryam Bar-Matthews, Gerhard W. Weber, Rolf Quam, Yossi Zaidner, Mina Weinstein-Evron

Our original claim, based on three independent numerical dating methods, of an age of ~185,000 years for the Misliya-1 modern human hemi-maxilla from Mount Carmel, Israel, is little affected by discounting uranium-series dating of adhering crusts. It confirms a much earlier out-of-Africa *Homo sapiens* expansion than previously suggested by the considerably younger (90,000 to 120,000 years) Skhul/Qafzeh hominins.

Full text: dx.doi.org/10.1126/science.aat8964

TECHNICAL COMMENT

PALEOANTHROPOLOGY

Comment on "The earliest modern humans outside Africa"

Warren D. Sharp[1]* and James B. Paces[2]

Hershkovitz *et al*. (Reports, 26 January 2018, p. 456) interpreted the Misliya-1 fossil maxilla as evidence of the earliest known anatomically modern human outside Africa. However, the fossil's reported age of 177,000 to 194,000 years relies on flawed interpretations of uranium-series data. We contend that those data support a minimum age of no more than ~60,000 to 70,000 years.

Misliya-1, part of the maxilla of an anatomically modern human from Misliya Cave, Israel, was assigned an age of 177 to 194 thousand years (ky) (*1*). Methods using U-series disequilibrium to date the fossil included combined U-series and electron spin resonance (US-ESR) on tooth enamel, laser-ablation ^{230}Th/U dating of dentine, and ^{230}Th/U dating of soil carbonate encrusting the maxilla. The US-ESR age estimate of 174 ± 20 ky (all errors 2σ) was considered a maximum age because the fossil was irradiated during three computed tomography scans prior to dating (*1*). Laser spots across a section of dentine yielded broadly consistent ^{230}Th/U age estimates between 83 ± 9 and 62 ± 3 ky. A ^{230}Th/U age estimate of 185 ± 8 ky on soil carbonate was interpreted as a minimum age for the fossil; however, that date is not supported by the U-series data in (*1*).

A reported age of 185 ± 8 ky for crust sample Maxilla-6 is crucial to the central claim of Hershkovitz *et al*., but that sample has clearly failed to maintain a closed U-series system. The authors erroneously reported an uncorrected age of "equilibrium" for Maxilla-6 [table S2 of (*1*)]. In fact, Maxilla-6 has an uncorrected ^{230}Th/^{234}U activity ratio greater than 1.0, indicating an excess of ^{230}Th over its parent ^{234}U (*2*). This condition is commonly produced when materials lose uranium relative to thorium after deposition, thereby violating a fundamental premise of radiometric dating. We contend that the reported age for Maxilla-6 was obtained by incorrectly computing detritus-corrected U-series isotope ratios, as described below.

Figure 1 shows measured and detritus-corrected activity ratios (solid ellipses and red crosses, respectively) for all eight maxilla crusts analyzed in (*1*). Maxilla-6 may have originally had a composition similar to other crusts; however, its measured ^{230}Th/^{238}U activity ratio is too high to be produced by closed-system radioactive decay, indicating that the sample has likely lost uranium after crust formation (Fig. 1, thick arrow).

Hershkovitz *et al*. recognized that measured U and Th compositions represent mixtures of authigenic soil carbonate and detrital silicate containing extraneous ^{230}Th, ^{234}U, and ^{238}U proportional to the amount of common Th (^{232}Th) present in the detritus. Using the observed ^{232}Th abundance and a model detritus composition, we derived the composition of the authigenic carbonate by mathematically subtracting U-series isotopes associated with the detritus from measured compositions. In Fig. 1, model detritus (red star) is assumed to have a Th/U value similar to Earth's average upper crust and U-series isotopic ratios in secular equilibrium (^{230}Th/^{238}U and ^{234}U/^{238}U activity ratios = 1.0) as in (*1*). Subtraction of detritus moves the calculated detritus-corrected compositions radially away from the model detritus by an amount that scales with the level of contamination (solid black arrows) (*3*). Proper correction for detritus in Maxilla-6 likewise moves its composition away from the model detritus, further into the open-system region (dashed arrow). However, the detritus-corrected composition reported for Maxilla-6 in (*1*) is displaced in the opposite direction (Fig. 1, red cross #6). The erroneous corrected composition is misleading because it implies that an age can be calculated for Maxilla-6, which has clearly

[1]Berkeley Geochronology Center, Berkeley, CA 94709, USA. [2]U.S. Geological Survey–Denver Radiogenic Isotope Lab, Box 25046, MS963, Denver Federal Center, Denver, CO 80225, USA. *Corresponding author. Email: wsharp@bgc.org

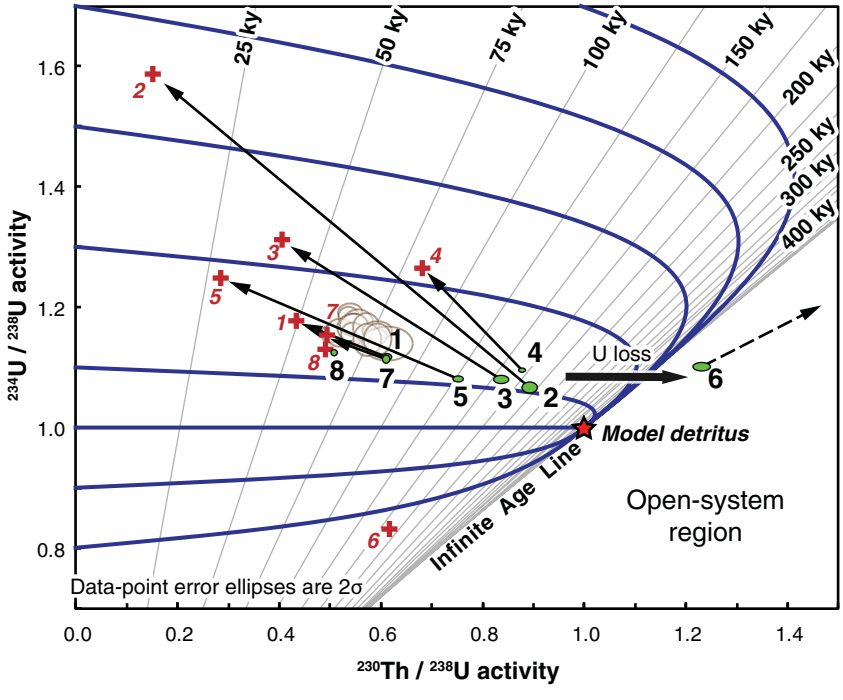


Fig. 1. Isotope evolution diagram showing U-Th activity ratios for crust samples Maxilla-1 through -8 calculated from data reported in (*1*). Straight lines are isochrons labeled in ky at their upper ends. Curved heavy lines show closed-system evolution with time for selected initial ^{234}U/^{238}U values. Uncorrected (green error ellipses) and detritus-corrected (red crosses) compositions of crusts are shown. Errors for detritus-corrected compositions are not shown for clarity but would be much larger than the ellipses shown for measured ratios. Thick arrow indicates expected displacement produced by loss of uranium (or gain of ^{230}Th). The region containing Maxilla-6, to the right of and below closely spaced isochrons defining the infinite-age line, cannot be reached by evolution in closed U-Th systems. Thin solid arrows that connect uncorrected and detritus-corrected compositions for crust samples project away from the model detritus (red star). Detritus-corrected compositions are those reported in (*1*), except for that of Maxilla-2, which was recalculated from uncorrected ratios in (*1*). Dashed arrow shows direction of accurate detritus correction for crust Maxilla-6. Laser-ablation data for dentine (open brown ellipses between 50- and 100-ky isochrons) are from table S1 of (*1*).

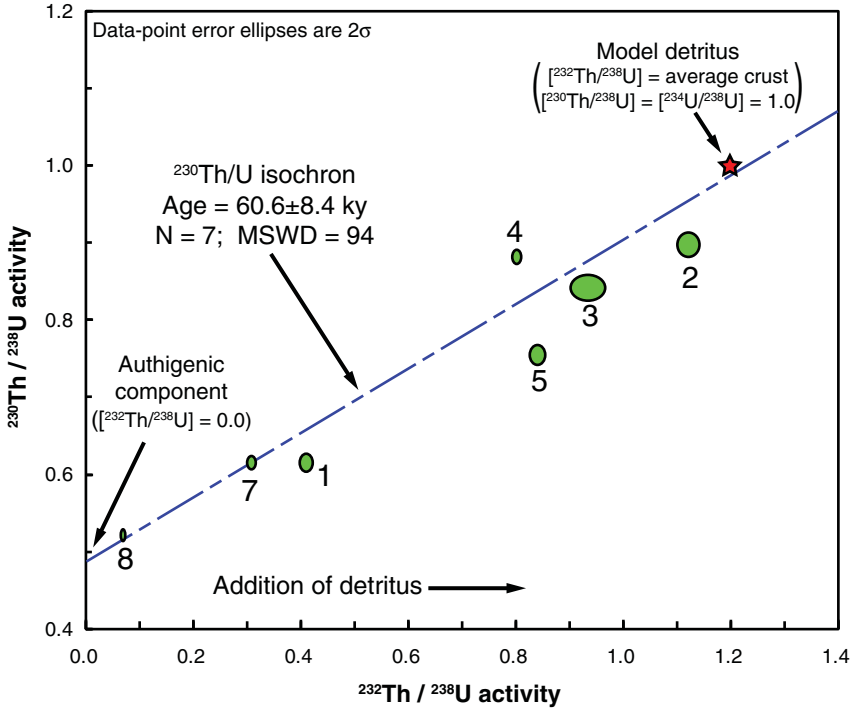


Fig. 2. Projection of a three-dimensional ^{230}Th-^{232}Th-^{234}U-^{238}U isochron showing activity ratios for all maxilla crusts analyzed in (*1*) except Maxilla-6. The isochron regression (dash-dot line) is controlled mainly by data for Maxilla-7 and -8, the least contaminated crusts, and is consistent with the assumed model detritus composition (red star). More contaminated crusts Maxilla-2 through -5 have ^{232}Th/^{238}U ratios approaching those of the model detritus. MSWD = mean square weighted deviates, a measure of the observed scatter relative to that expected from analytical errors alone.

failed to maintain a closed U-series system. Moreover, the inaccurate corrected composition for Maxilla-6 in (*1*) yields an unreasonable initial ^{234}U/^{238}U activity of ~0.75, unlike other soil carbonates from Misliya Cave and elsewhere that are enriched rather than depleted in ^{234}U relative to ^{238}U (*4*, *5*).

Calculations using Isoplot (*6*), an application widely used to reduce U-series data, produce an inaccurate detritus-corrected composition and age for Maxilla-6 similar to those given in (*1*). This is apparently a consequence of the unusual composition of Maxilla-6, which has a higher Th/U value than the model detritus, making it unsuitable for U-series dating. We infer that the ^{232}Th-rich composition of Maxilla-6 lies outside the range of applicability of the algorithm used by Isoplot for calculating detritus-corrected compositions.

Detritus-corrected ^{230}Th/U age estimates of 70 to 19 ky were reported for seven additional maxilla crust samples (*1*). All are moderately to highly contaminated by silicate detritus. Assuming a detritus composition like that in (*1*) and uncertainty propagation as in (*5*, *7*), we calculate ^{230}Th/U ages of 60 ± 20 and 61.7 ± 3.8 ky for the two least contaminated samples, Maxilla-7 and -8. Use of a three-dimensional ^{230}Th-^{232}Th-^{234}U-^{238}U isochron approach, which requires no assumptions regarding the isotopic composition of detritus, shows that Maxilla-7 and -8 are approximately collinear with the model detritus composition (Fig. 2). A regression using all measured data, save Maxilla-6, yields an estimated isochron age of 60.6 ± 8.4 ky, albeit with a high degree of scatter. Finite age estimates can be calculated for the more severely contaminated samples Maxilla-1 through -5; however, our results yield age uncertainties ranging from 56 to >350%. Much smaller errors for corrected U-series ages are reported in Hershkovitz *et al.* [table S2 of (*1*)] as a result of failing to propagate uncertainty arising from the large detritus corrections. We contend that such errors severely underestimate actual uncertainties.

We note that compositions and age estimates of the most suitable crusts, Maxilla-7 and -8, are broadly consistent with results from laser-ablation U-series analyses of fossil dentine (Fig. 1, open ellipses) with a mean age of 70.2 ± 1.6 ky (*1*).

In sum, most samples of carbonate encrusting Misliya-1 analyzed in Hershkovitz *et al.* are unsuitable for U-series age interpretation. Only two carbonate crust samples are sufficiently pure to provide reliable U-series dates. If they have maintained closed U-series systems, results for Maxilla-7 and -8 support a minimum age of ~60 ky for the fossil, similar to values determined by laser ablation on dentine (*1*). We contend that there are no reliable U-series dates for Misliya-1 older than ~70 ky. Previously known fossils of anatomically modern humans in the Levant are as old as 90 to 120 ky (*8*–*10*). Thus, the claim that Misliya-1 is the earliest such fossil is not supported by U-series dating in Hershkovitz *et al.* Published thermoluminescence dates from burnt lithic artifacts from Misliya Cave suggest ages of ~140 to 212 ky for associated sediments (*1*, *11*), but the available dates on the human fossil cannot rule out the possibility that the fossil dates to a younger interval.

REFERENCES AND NOTES

1. I. Hershkovitz *et al.*, *Science* **359**, 456–459 (2018).
2. The decay rate or activity of a nuclide is given by λn, where λ is the decay constant and n is the number of atoms. U-series nuclides in a material that remains undisturbed for a few million years reach a state of secular equilibrium where their activities become equal and their activity ratios reach unity.
3. We were unable to reproduce the detritus-corrected composition of crust sample Maxilla-2 from the uncorrected ratios in table S2 of (*1*); therefore, we show our result for the detritus-corrected composition of Maxilla-2 in Fig. 1.
4. E. J. Oerter *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 919–924 (2016).
5. K. R. Ludwig, J. B. Paces, *Geochim. Cosmochim. Acta* **66**, 487–506 (2002).
6. K. R. Ludwig, *Berkeley Geochronological Center Special Publication No. 5* (2012).
7. Assumed activity ratios of detritus: ^{232}Th/^{238}U = 1.2 ± 0.6, ^{230}Th/^{238}U = 1.0 ± 0.25, and ^{234}U/^{238}U = 1.0 ± 0.25. Decay constants used are those in (*12*).
8. N. Mercier *et al.*, *J. Archaeol. Sci.* **20**, 169–174 (1993).
9. C. B. Stringer, R. Grün, H. P. Schwarcz, P. Goldberg, *Nature* **338**, 756–758 (1989).
10. H. P. Schwarcz *et al.*, *J. Hum. Evol.* **17**, 733–737 (1988).
11. H. Valladas *et al.*, *J. Hum. Evol.* **65**, 585–593 (2013).
12. H. Cheng *et al.*, *Earth Planet. Sci. Lett.* **371–372**, 82–91 (2013).

13 April 2018; accepted 29 August 2018
10.1126/science.aat6598

TECHNICAL RESPONSE

PALEOANTHROPOLOGY

Response to Comment on "The earliest modern humans outside Africa"

Israel Hershkovitz[1,2]*†, Mathieu Duval[3,4]†, Rainer Grün[3,5], Norbert Mercier[6], Helene Valladas[7], Avner Ayalon[8], Miryam Bar-Matthews[8], Gerhard W. Weber[9], Rolf Quam[10], Yossi Zaidner[11], Mina Weinstein-Evron[11]

Our original claim, based on three independent numerical dating methods, of an age of ~185,000 years for the Misliya-1 modern human hemi-maxilla from Mount Carmel, Israel, is little affected by discounting uranium-series dating of adhering crusts. It confirms a much earlier out-of-Africa *Homo sapiens* expansion than previously suggested by the considerably younger (90,000 to 120,000 years) Skhul/Qafzeh hominins.

In reply to Sharp and Paces' comments on our recent paper (*1*), we here clarify some points that may have remained unappreciated in our initial work. Sharp and Paces (*2*) claim that Misliya-1 may not be as old [177 to 194 thousand years (ka)] as we argued (*1*). They propose an alternative interpretation for the U-series data we reported on calcitic crusts [table S1 of (*1*)], leading to the calculation of an apparent age of 60 to 70 ka. This age provides a minimum age for Misliya-1, as we also noted in the original paper (*1*). Although their reanalysis appears robust, Sharp and Paces give little credence to the importance of other dating evidence, such as the thermoluminescence (TL) dating of 23 samples from the associated Early Middle Paleolithic (EMP) lithic assemblage of the site (*3*). In this reply, we address (i) the uncertainties and limitations in the isochron U-series analysis of the calcitic crusts, (ii) the comprehensive evidence presented in the original paper for the specimen's antiquity, (iii) the unlikely possibility of later intrusion, (iv) the robust TL dating evidence, and (v) the inappropriate use of the Skhul and Qafzeh hominins to support their claim.

The Misliya-1 crusts were drilled in several locations from adhering lithified sediments above the molars (Fig. 1). The U-series isochron approach requires contemporaneous samples. However, there is layering within the crusts, which shows that they were formed during various wet phases over a longer period of time. Therefore, an isochron approach seems inappropriate. Additionally, although mentioned by Sharp and Paces (but later ignored), the scatter of the points around their regression line is quite large (13%), which may actually indicate a higher true age uncertainty than that reported in their comment.

Sharp and Paces ignore several lines of evidence, all pointing to a similar time range for Misliya-1:

1) All six EMP units in Misliya Cave contain only an Early Levantine Mousterian lithic industry (Tabun D-type). This industry is dated 276 to 140 ka (Fig. 2) on the basis of radiometric data obtained from Misliya (*3*), Hayonim (*4*), and Tabun (*5*).

2) The TL dating of nine burned flints from the upper part of the EMP archaeological layer from squares close to Misliya-1 (N-12, L-10) (Fig. 2) yielded a mean age of 179 ± 48 ka [all errors in this reply are 2σ; raw data from (*3*)]. The somewhat distant square (J-15) exhibited the same TL age range (Fig. 2). The combined mean age for these samples is 185 ± 50 ka (n = 15).

3) Direct dating of Misliya-1 provided a U-series age of 70.2 ± 1.6 ka and a combined U-series and electron spin resonance (US-ESR) age of 174 ± 20 ka. Because uranium uptake may be delayed after the death of the organism, and because it is difficult to accurately evaluate the radiation effects from previous computed tomography (CT) scanning, these two ages should be regarded as the minimum and maximum age brackets for the fossil, respectively. Misliya-1 was CT-scanned three times prior to the ESR dating analysis (*1*). The x-ray dose during CT scanning is highly variable (*6*–*8*). On the basis of our recent study (*7*), we could roughly estimate that the total x-ray dose was approximately 30 Gy, resulting in an effective equivalent dose (D_E) value of 98.3 ± 5.3 Gy. The corresponding US-ESR age would thus be 152 ± 24 ka (i.e., agreeing within error with the TL results). However, considering the significant uncertainty in the true

Fig. 1. The Misliya-1 specimen from a lateral view. The crust under the zygomatic arch was analyzed for U-Th dating. Note the different layers of the crust and the embedded foreign elements (bones and flints) suggesting different depositional episodes.

[1]Department of Anatomy and Anthropology, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 6997801, Israel. [2]Dan David Center for Human Evolution and Biohistory Research and Shmunis Family Anthropology Institute, Steinhardt Museum of Natural History, Tel Aviv University, Tel Aviv 6997801, Israel. [3]Australian Research Centre for Human Evolution (ARCHE), Environmental Futures Research Institute, Griffith University, Nathan, QLD 4111, Australia. [4]Centro Nacional de Investigación sobre la Evolución Humana (CENIEH), 09002 Burgos, Spain. [5]Research School of Earth Sciences, Australian National University, Canberra, ACT 2601, Australia. [6]Institut de Recherche sur les Archéomatériaux, UMR 5060 CNRS–Université Bordeaux-Montaigne, Centre de Recherche en Physique Appliquée à l'Archéologie (CRP2A), Maison de l'archéologie, 33607 PESSAC Cedex, France. [7]LSCE/IPSL, UMR CEA-CNRS-UVSQ, 91198 Gif-sur-Yvette Cedex, France. [8]Geological Survey of Israel, Jerusalem 95501, Israel. [9]Department of Anthropology and Core Facility for Micro-Computed Tomography, University of Vienna, A-1090 Vienna, Austria. [10]Department of Anthropology, Binghamton University, Binghamton, NY 13902, USA. [11]Zinman Institute of Archaeology, University of Haifa, Mount Carmel 3498838, Israel.
***Corresponding author. Email: anatom2@post.tau.ac.il**
†These authors contributed equally to this work.

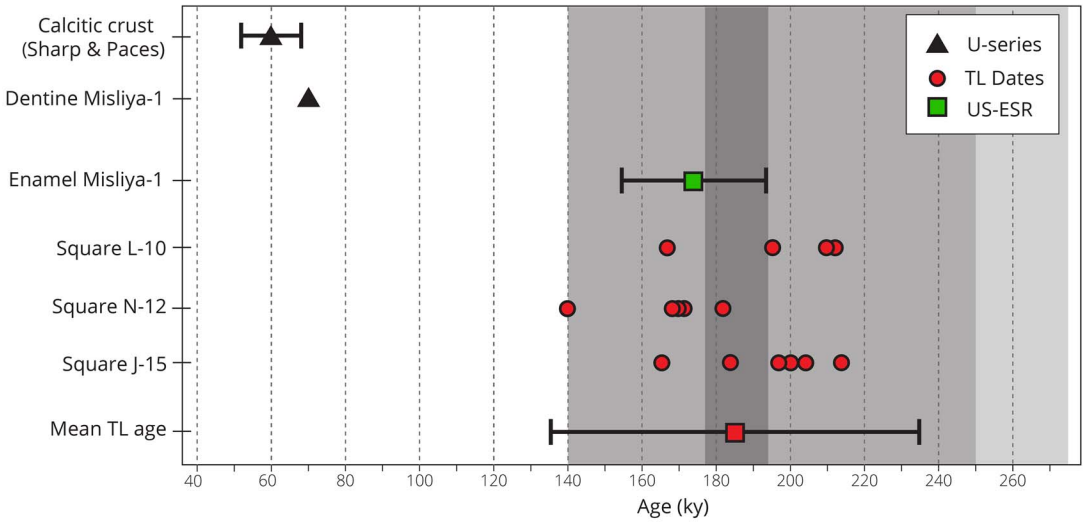


Fig. 2. Overview of the dating results. All errors are 2σ. For clarity, the individual errors on the TL ages (red dots) are not displayed. The combined US-ESR age (green square) should be regarded as a maximum age estimate for Misliya-1 (based on this method). The U-series age carried out on the dentine of the Misliya-1 lateral incisor should be considered as a minimum age estimate of the specimen. The TL ages from Misliya Cave (red dots) from the various relevant squares fall within the boundaries of the EMP period (entire gray-shaded area), as previously determined on the basis of TL data from Tabun Cave and Hayonim Cave (140 to 276 ka) [data from Mercier *et al*. (*4*) and Mercier and Valladas (*5*)]. Intermediate gray marks the EMP boundaries at Misliya Cave (140 to 250 ka) as published by Valladas *et al*. (*3*). The dark gray band marks the dating boundaries for Misliya-1 as published in Hershkovitz *et al*. (*1*) (177 to 194 ka).

x-ray dose absorbed by the tooth sample, regarding the cited date of 174 ± 20 ka as a maximum age is the most straightforward and reasonable interpretation of the combined US-ESR result.

4) The major criticism of Sharp and Paces relates to the U-series dates on the calcitic crusts, specifically to sample #6. The U-series dates were added at the request of one of the reviewers. We were aware of and mentioned the high detrital thorium in these samples and noted that corrections had a large impact on the apparent U-series ages [see table S1 of (*1*)]. The U-series results were therefore presented as additional support for the dates obtained by the other methods. It is worth noting that a crust deposited on a flint from a nearby square from a depth close to that of the maxilla yielded an apparent U-series date similar to sample #6 [table S1 of (*1*), sample #10, ~172 ka]. This sample also had a relatively high detrital thorium content. Even if the original U-series dates are discounted, there is no real impact on the dating of Misliya-1 (Fig. 2).

Sharp and Paces hint at the possibility that Misliya-1 could be much younger than proposed, within the range of the Skhul/Qafzeh fossils (90 to 120 ka). Because the EMP context cannot be younger than 140 ka (see above), this would imply a later intrusion. This was discussed and dismissed (*1*) because there was no evidence for any culture later than the EMP at the site, which was apparently abandoned once the roof of the cave had collapsed.

Misliya-1 was derived from within a clear archaeological context, rich in lithics, animal bones, and well-defined in situ archaeological features (hearths). Detailed taphonomic and geoarchaeological studies (*9*, *10*) also indicate excellent preservation of original features (e.g., plant bedding) and no substantial postdepositional transport. This all suggests that the archaeological layers and the human fossil are coeval.

Sharp and Paces briefly mention the TL ages from Misliya but downplay their importance. The samples closest to Misliya-1 yielded a mean age of 179 ± 48 ka and, combined with other samples in the vicinity, a mean age of 185 ± 50 ka (see above). EMP lithic assemblages were found in two other caves. At Tabun (*5*), three samples from unit IX produced a mean age of 256 ± 52 ka, and two spatially isolated units yielded mean ages of 222 ± 54 ka (n = 4) and 196 ± 42 ka (n = 3). At Hayonim (*4*), layers F and Lower E produced 77 TL ages ranging from 140 ± 32 ka to 251 ± 40 ka. None of the three sites contains any EMP lithic industry younger than 140 ka. As there is no evidence to suggest that Misliya-1 is intrusive, we reasonably conclude that the mean TL age of 185 ± 50 ka provides a reliable indirect age estimate for this specimen.

Sharp and Paces present the ages of the Skhul and Qafzeh hominins (90 to 120 ka) as supportive evidence for a possible younger age for Misliya-1. The EMP of Misliya Cave, with its abundant blades and typical long Levallois and non-Levallois points (*11*), is distinctly different and stratigraphically older than the Middle Paleolithic found at Skhul and Qafzeh. The fact that there are other hominins dated to other time ranges in the area has little bearing on the age of Misliya-1.

The study of Sharp and Paces does not contradict our chronology for Misliya-1 (*1*). Although their reinterpretation of the apparent U-series ages on the calcitic crusts may be sound, the resulting minimum age of ~60 to 70 ka is neither new (it was already discussed in the original paper) nor incompatible with the other chronological data available (Fig. 2). Several lines of evidence, including direct dating of the specimen, the EMP lithic industry, and the TL dates, clearly support an early date for Misliya-1 of at least 140 ka, and most likely around 185 ka, consistent with our original conclusions.

REFERENCES AND NOTES

1. I. Hershkovitz *et al*., *Science* **359**, 456–459 (2018).
2. W. D. Sharp, J. B. Paces, *Science* **362**, eaat6598 (2018).
3. H. Valladas *et al*., *J. Hum. Evol.* **65**, 585–593 (2013).
4. N. Mercier *et al*., *J. Archaeol. Sci.* **34**, 1064–1077 (2007).
5. N. Mercier, H. Valladas, *J. Hum. Evol.* **45**, 401–409 (2003).
6. R. Grün, S. Athreya, R. Raj, R. Patnaik, *Archaeol. Anthropol. Sci.* **4**, 25–28 (2012).
7. M. Duval *et al*., *Quat. Geochronol.* **47**, 120–137 (2018).
8. M. Duval, L. Martín-Francés, *Am. J. Phys. Anthropol.* **163**, 205–212 (2017).
9. M. Weinstein-Evron *et al*., *Paleoanthropology* **2012**, 202–228 (2012).
10. R. Yeshurun, G. Bar-Oz, M. Weinstein-Evron, *J. Hum. Evol.* **53**, 656–677 (2007).
11. Y. Zaidner, M. Weinstein-Evron, *Before Farming* **2012**, 1–23 (2012).

ACKNOWLEDGMENTS

M.D.'s research is currently funded by the Australian Research Council Future Fellowship grant FT150100215.

8 May 2018; accepted 4 September 2018
10.1126/science.aat8964

S&T Policy Fellowships find a home in the judiciary

Fellows work on independent research topics during their year at the Federal Judicial Center

By **Becky Ham**

While the American Association for the Advancement of Science's Science & Technology Policy Fellowships are more than four decades old, the program has been sending scientists to share their expertise with the U.S. judiciary for only 4 years.

Since 2014, the program has placed four fellows at the Federal Judicial Center in Washington. Their presence is small compared to the hundreds of fellows deployed to the federal legislative and executive branches, but the judicial fellows already have made a significant impact, said James Eaglin, who directs the center's research division and serves as mentor to the policy fellows.

The fellows' contributions to the federal courts "have proved invaluable," Eaglin said. "When questions come up from a colleague that may relate to the intersection of law and science, we have the fellows as a recognized source of expertise and a resource for us to help resolve or answer those questions."

The center conducts and promotes research on federal judicial procedures and administration and provides continuing education and training for judges and federal court employees. Because of its engagement in a wide range of issues, many of which are relevant to science and technology, fellows at the center often have an opportunity to work in areas that correspond with their interests and expertise.

> ***"...science and law are the two professions on which our modern society is fundamentally based."***
>
> **Honorable Barbara Rothstein,** senior U.S. district judge

In addition to the scientific or engineering credentials required of all Science & Technology Policy fellows, the judicial fellowship requires at least 3 years of professional work experience. A law degree or other legal expertise is preferred but not mandatory.

Shubha Ghosh, who has a Ph.D. in economics from the University of Michigan and a J.D. from Stanford Law School, was the program's inaugural judicial fellow. He came to the program in mid-career, after several years working as a professor and consultant. At the center, he researched attorneys' fees in patent litigation and patent law related to DNA research. The year in Washington "enhanced my scholarly focus," he said, leading him to write several journal articles and book chapters on biotechnology patenting. Now a professor at Syracuse University College of Law, Ghosh said that the fellows' experience "has also informed my teaching, and pushed me to want to teach more about policy in the future."

By contrast, the 2017–18 judicial fellow, neuroscientist Andrea Gaede, applied for the fellowship as an early-career scientist with no legal background. She had developed an interest in neuroethics during postdoctoral work at the University of British Columbia and the University of Alberta. She worked with staff in the center's education division and others to develop interactive online training modules for federal judges and staff on the basics of recent neuroscience research and its use in the legal system.

PHOTO: B CHRISTOPHER/ALAMY STOCK PHOTO

Fellows work at the Thurgood Marshall Federal Judiciary Building in Washington, D.C.

These modules cover topics from addiction science to brain imaging technologies to how differences in levels of brain development impact behavior. One of the challenges in creating the modules was finding the right level for the training, Gaede said. "Some judges may not have had a biology class since high school, and some of them might know what p-hacking [a bias in data collection and analysis] is," she explained. "Our product has to deal with a real range of skill and background."

Since his 2016–17 fellowship, Duke University researcher Pate Skene has worked with others on amicus briefs for cases involving standards for forensic evidence and the effects of family separations and stress on brain development in children and teens.

Mamadi Corra, a sociology professor from East Carolina University, began his fellowship year in September. His research project will focus on physical barriers to accessing court data through online systems such as PACER (Public Access to Court Electronic Records).

"We are grateful for the support we receive from the Gordon and Betty Moore Foundation to be able to offer the judicial fellowship, and we are eager to be able to bring more fellows to the federal judiciary in the future," said Jennifer Pearl, director of the fellowship program.

The fellows provide expertise that increasingly is in demand by the federal courts, as judges and their colleagues grapple with the legal implications of rapidly advancing technology and sort out the validity of scientific evidence and expert testimony, said the Honorable Barbara Rothstein, senior U.S. district judge for the District of Columbia.

Rothstein and other speakers, including American University Washington College of Law constitutional law professor Stephen Wermiel, shared a broad overview of the judicial branch with the 2018–19 fellowship class during their two-week orientation held in Washington in September. The cultures and methods of science and the law differ, Rothstein said, especially in matters of how long it takes to resolve a dispute and what kind of consensus must be built around evidence.

"But science and law are the two professions on which our modern society is fundamentally based," she said. "At a time when there are alternative facts, false news, both science and the law stand as bulwarks in the search for truth."

AAAS Council reminder

The next meeting of the AAAS Council will take place during the 2019 AAAS Annual Meeting in Washington, D.C., and will begin at 9:00 a.m. on 17 February 2019 at the Omni Shoreham Hotel.

Individuals or organizations wishing to present proposals or resolutions for possible consideration by the council should submit them in written form to AAAS Chief Executive Officer Rush Holt by 10 January 2019. This will allow time for them to be considered by the Committee on Council Affairs at its winter meeting.

Items should be consistent with AAAS's objectives and be appropriate for consideration by the council. Resolutions should be in the traditional format, beginning with "Whereas" statements and ending with "Therefore be it resolved."

Late proposals or resolutions delivered to the AAAS chief executive officer in advance of the February 2019 open hearing of the Committee on Council Affairs will be considered, provided that they deal with urgent matters and are accompanied by a written explanation of why they were not submitted by the 10 January deadline. The Committee on Council Affairs will hold its open hearing at 2:30 p.m. on 16 February 2019 in the Omni Shoreham Hotel.

Mass Media fellows wrap up productive summer

Smaller, regional media outlets in the United States were the focus of the 2018 program

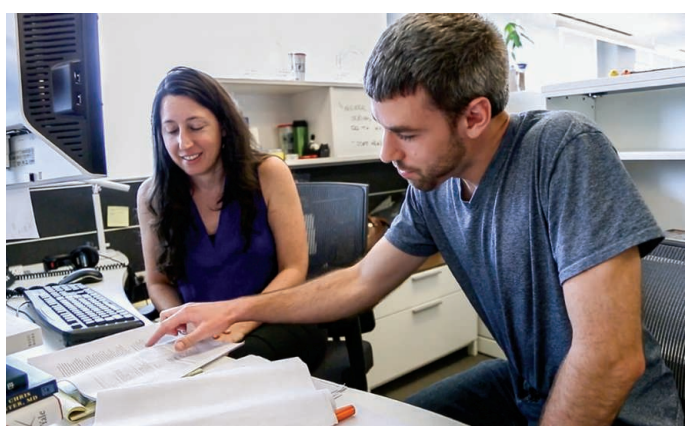
By **Becky Ham**

Kevin Davenport is working toward a Ph.D. in experimental physics, but he's recently become a bit of an expert on long-billed curlews. During his summer as an American Association for the Advancement of Science Mass Media fellow at the *Idaho Stateman*, Davenport wrote a story about the unusually high level of human threats to the migratory birds in southwest Idaho.

"I got a lot of immediate feedback from the community," said Davenport of the story, "which is one of the great things about working for a local paper."

The AAAS Mass Media Fellows program pairs scientists and engineers like Davenport with media outlets in the United States, giving the researchers a chance to report and write news stories for 10 weeks each summer. Now in its 44th year, the program has supported more than 750 fellows, including highly regarded scientists and journalists.

This year, the program made "major efforts" to place fellows at media outlets away from the U.S. coasts, said Shirley Malcom, director of AAAS Education and Human Resources. At an event held 20 August to showcase the fellows' work, Malcom said, "As I walked around I got excited, because I saw a lot of science being reported in the middle of the country."



Mass Media fellow Paul Chisholm works with NPR editor Carmel Wroth.

This was the first year that the *Statesman* has worked with a Mass Media fellow, said the paper's community engagement editor, Bill Manny. Davenport came to work eager to learn, Manny said, and the two of them had a "Socratic summer" of talking over how to write stories with science in them—as opposed to "science stories," Manny recalled.

PHOTO: NEIL ORMAN/AAAS

Another of Davenport's articles explained fire damage to a concrete overpass that was the result of a fatal interstate accident in west Boise. "That was the story that got the most readership of all the stories he did for us," said Manny, "and I think it was where he really figured out how you can use a science perspective and a science background on just good basic journalism."

Several of this year's fellows said their science background had prepared them in unexpected ways for their summer jobs. Katie Wu, a Harvard microbiology Ph.D. student who worked with *Smithsonian* magazine online, found that her time as a scientist "has made me a much better writer," she said. "I'm not afraid of failure, and I'm not afraid to ask what others might perceive as stupid questions."

When Alejandra Borunda finished her master's degree in journalism, "I don't think I had the confidence to tell the kinds of science stories that I wanted to tell," she said. Borunda, who spent her summer working at *National Geographic*, is now studying for a Ph.D. in earth and environmental sciences at Columbia University. "It took me going back to graduate school and really immersing myself in the nitty gritty of science to realize that I really did have that capability."

Other fellows said that they were challenged and excited by the pace and environment of the newsroom. "I'm learning a lot more than at any other point in my life, and that includes grad school," said Paul Chisholm, a recent Ph.D. graduate in entomology from Washington State University who worked at the National Public Radio science desk. "I kind of compare it to a chunk of carbon that is being put under intense pressure. It's a lot of heat, a lot of pressure, but then that piece of carbon comes out the other end as a diamond."

Kathryn Furby, a recent Ph.D. graduate in marine biology from the University of California, San Diego, wrote about topics from whale sharks to liver disease for *The Washington Post*. She said that the best part of each week at the paper was the team meeting where reporters and editors discussed their assignments. "I feel like journalists are always gathering stories as a part of their jobs, and that makes them really interesting colleagues."

As in previous years, some of the fellows hope to use their summer experience to launch a new career in journalism, while others want to take the lessons they've learned back to their work in academia. Francisco Guerrero Bolano, who is finishing a dual Ph.D. in forestry and water resources science at Oregon State University, said that he wants to become more involved in training scientists to become better communicators after his time as a fellow at CNN Español. "Something that I learned from my experience at CNN is that we can compete with entertainment, and that we have an audience, and that audience is sitting in our classrooms and in conferences."

Anna Groves, who worked for the *Milwaukee Journal Sentinel* this summer, received her Ph.D. in plant biology from Michigan State University in 2018. She was hired as an assistant editor at *Discover* magazine after her fellowship.

One of her favorite stories from this summer covered drunk driving and breathalyzer tests. "This story was the one that most of my friends were like, 'I love that story. I didn't know how that worked, but now I do,'" she said. "I think that's a really fun takeaway, for someone to read something and actually learn something that they will remember."

RESEARCH

Soft magnetic materials reviewed

Silveyra et al., *p. 418*

IN *SCIENCE* JOURNALS

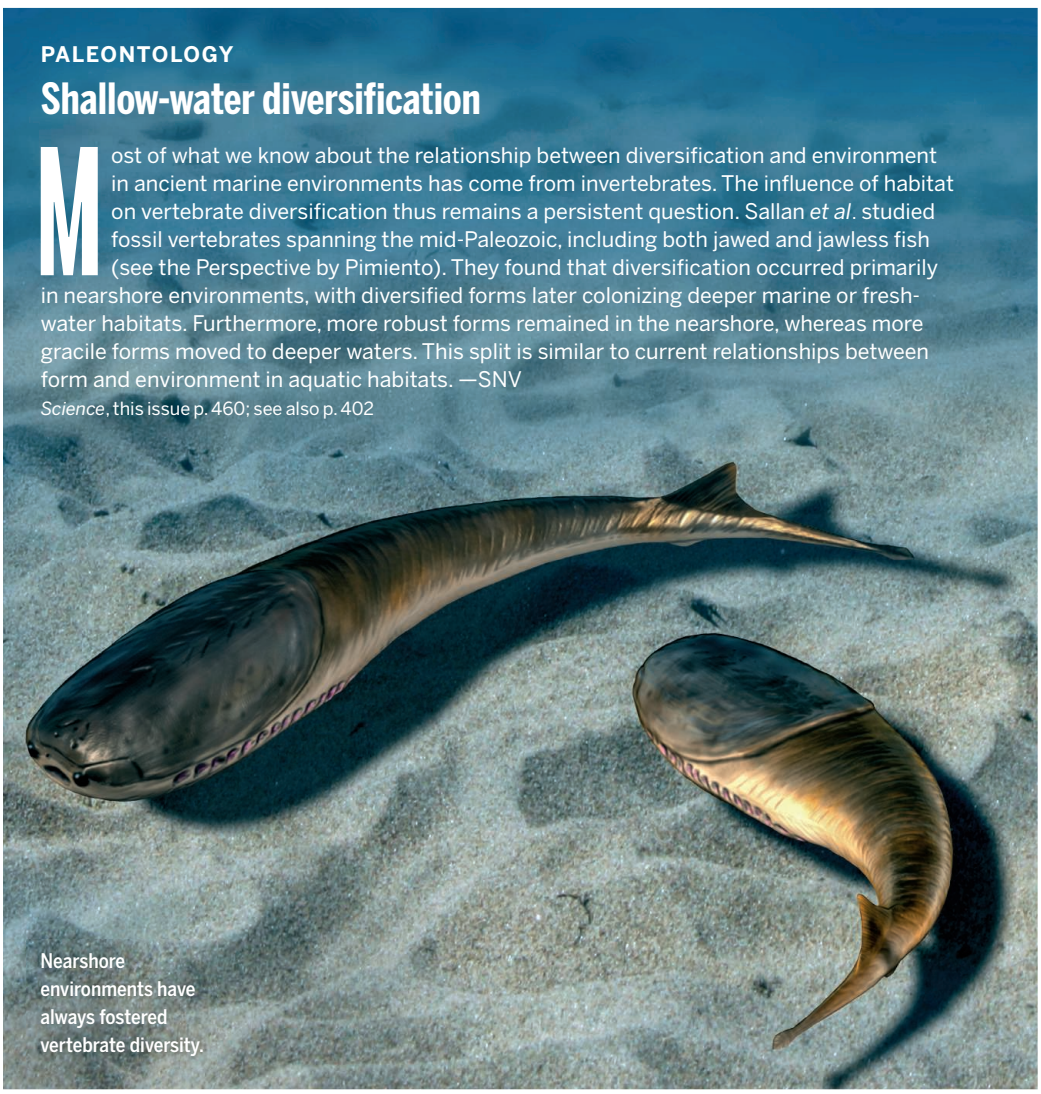
Edited by **Stella Hurtley**

PALEONTOLOGY

Shallow-water diversification

Most of what we know about the relationship between diversification and environment in ancient marine environments has come from invertebrates. The influence of habitat on vertebrate diversification thus remains a persistent question. Sallan *et al*. studied fossil vertebrates spanning the mid-Paleozoic, including both jawed and jawless fish (see the Perspective by Pimiento). They found that diversification occurred primarily in nearshore environments, with diversified forms later colonizing deeper marine or freshwater habitats. Furthermore, more robust forms remained in the nearshore, whereas more gracile forms moved to deeper waters. This split is similar to current relationships between form and environment in aquatic habitats. —SNV

Science, this issue p. 460; see also p. 402



Nearshore environments have always fostered vertebrate diversity.

CREDITS (FROM TOP): SILVEYRA *ET AL.*; SALLAN *ET AL.*

SOLAR CELLS

Staying in the black phase

Hybrid perovskite solar cells often use the more thermally stable formamidinium (FA) cation rather than methylammonium, but its larger size can create lattice distortion that results in an inactive yellow phase. Turren-Cruz *et al*. show that by using iodide instead of bromide as the anion (to create a redder bandgap) and an optical mix of cesium, rubidium, and FA cations, they can make solar cells with a stabilized efficiency of more than 20%. No heating steps above 100°C were needed to create the preferred black phase. —PDS

Science, this issue p. 449

NEUROSCIENCE

A close view of the paraventricular thalamus

The paraventricular thalamus is a relay station connecting brainstem and hypothalamic signals that represent internal states with the limbic forebrain that performs associative functions in emotional contexts. Zhu *et al*. found that paraventricular thalamic neurons represent multiple salient features of sensory stimuli, including reward, aversiveness, novelty, and surprise. The nucleus thus provides context-dependent salience encoding. The thalamus gates sensory information and contributes to the sleep-wake cycle through its interactions with the cerebral cortex. Ren *et al*. recorded from neurons in the paraventricular thalamus and observed that both population and single-neuron activity were tightly coupled with wakefulness. —PRS

Science, this issue p. 423, p. 429

MICROBIOTA

Transmission of the gut community

Natural transmission of the mammalian microbiota is poorly understood. Some genera of bacteria are transmitted from mothers to offspring, whereas others are acquired from the wider environment. Moeller *et al.* derived inbred mouse lines from two wild populations of mice with distinct microbiota and monitored the populations' microbiomes for 3 years while they were kept in the same animal facility. The microbiota of the two mouse lineages remained distinct even after 10 generations. Most microbiota genera transmitted vertically. Those taxa that transmitted horizontally through the shared environment of the animal facility tended to be those that include pathogens. —CA

Science, this issue p. 453

Published by AAAS

IN OTHER JOURNALS

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

NEUROSCIENCE

Heart rate and blood pressure control

PIEZO1 and PIEZO2 are two mechanically activated ion channels that are highly expressed in lungs, bladder, and skin. Zeng *et al*. found that both ion channels are expressed in sensory neurons of a ganglion complex that contribute to the baroreflex, a homeostatic mechanism that helps to keep blood pressure stable (see the Perspective by Ehmke). Conditional double knockout of PIEZO1 and PIEZO2 in these neurons abolished the baroreflex and disrupted blood pressure regulation and heart rates in mice. These changes were very similar to those seen in patients with baroreflex failure. In mice, selective activation of PIEZO2-expressing ganglion neurons triggered immediate increases in heart rate and blood pressure. —PRS

Science, this issue p. 464; see also p. 398

ARCHAEOLOGY

Pre-Clovis projectile points in North America

The Clovis culture was a prehistoric Paleoindian culture. Archaeological research over the past decades has refuted the Clovis-first hypothesis in the peopling of the Americas. However, controversy continues over the timing of earlier migrations, the identification and description of the material culture possessed by these early migrants, and the likely routes by which these peoples entered North America. Waters *et al*. studied the Debra L. Friedkin site in Texas to resolve some of these questions. They performed careful stratigraphic excavation and a meticulous program of chronometric dating to identify a style of projectile points with triangular blades and stemmed bases. These projectiles were dated to have been made between 13,500 and 15,500 years ago and were found below typical Clovis-style points dated to 13,000 years ago. Thus, either the stemmed form was transformed locally over time into the more widespread Clovis style or it is a style created by people of an earlier and separate migration. —MSA

Sci. Adv. 10.1126/sciadv.aat4505 (2018).

OPTICS

Shrinking synchrotron generation

The generation of synchrotron radiation is typically achieved by accelerating charges in large magnetic fields. Synchrotron facilities are usually the realm of large international organizations. Henstridge *et al*. show that the interaction of a femtosecond light pulse moving in an arc on a specially designed metasurface can also generate synchrotron radiation. In this case, the synchrotron radiation at terahertz frequencies was produced by the nonlinear polarization induced by the light pulse. The results hold promise for the development of powerful on-chip light sources. —ISO

Science, this issue p. 439

CANCER

Killing tumors by targeting their neighbors

Pancreatic cancer is infamous for its bad prognosis as well as for its dense stroma. Most therapies target the tumor cells themselves rather than the stroma. Zhou *et al*. identified a therapeutic target called DKK3, which is produced by pancreatic stellate cells. This protein was present in the majority of human pancreatic tumors sampled. Ablating DKK3, either by genetic means or with a monoclonal antibody, provided a potentially effective treatment. Antibody treatment reduced tumor growth and extended survival in mouse models, especially when combined with an immune checkpoint inhibitor. —YN

Sci. Transl. Med. **10**, eaat3487 (2018).

ELECTROCHEMISTRY

Coarsening platinum electrodes

Platinum (Pt) electrodes can undergo oxidation and reduction that can alter their surface morphology. Deng *et al*. used atomic force microscopy to follow changes in surface morphology of a smooth Pt foil in sulfuric acid as it was cycled between various oxidizing and reducing potentials. After oxidizing the surface at or above 1.8 volts, mild reduction (~0.8 volts) created Pt ions in the surface region, which, upon further reduction, deposited Pt nanoparticles and created a rough surface. The formation of nanoparticles depended critically on oxide thickness. Thinner oxides created fewer nucleation sites, which led to fewer, but larger, nanoparticles. —PDS

J. Am. Chem. Soc. **140**, 13285 (2018).

EDUCATION

Can you recognize recognition?

Research addressing the gender imbalance in physics highlights the importance of a physics teacher recognizing a female student as being a "physics person." But what does this recognition look like? To examine this question, Hazari and Cass collected video recordings, field notes, interviews, and surveys from high-school physics teachers and students and developed a case study of a student who, because of her teacher's modes of recognition, began to see herself as a physics person. Results show that although the teacher did not perceive the student as a physics person, the student felt another way based on her perception of his actions (she believed her teacher saw her as a physics person), suggesting that regardless of what teachers themselves believe, their behaviors can enable students' physics identity development. —MMc

Phys. Teach. **56**, 442 (2018).

ENVIRONMENTALISM

Measuring charity

How do we know that a donation made to a charity is effective? We can total up the funds raised by a civil society group, but this measure misses if, and how well, the money translates into public good. Water pollution in the United States is still a problem, despite the 1972 Clean Water Act. Grant and Langpap examined the

PHOTO: ARTERRA/GETTY IMAGES

ECOLOGY

Pollination in an agricultural landscape

Plants and their insect pollinators form ecological networks across landscapes, networks that can be altered by human activity, principally agriculture. Redhead *et al*. assessed the composition and structure of plant-pollinator networks across Great Britain, assembling citizen-science records of all plant-pollinator interactions. By simulating extinction of plant species from communities, they were then able to assess the relative robustness of these networks to changes in community composition. They found that networks in highly agricultural landscapes are more robust to perturbation, as the pollinators in these landscapes tend to be generalists, and the plants less extinction prone. Nevertheless, they note that managing landscapes for crop pollination may not be a recipe for wider biodiversity conservation. —AMS

Ecol. Lett. 10.1111/ele.13157 (2018).

Italian honeybee, a generalist pollinator, visiting chive flowers

accounts of more than 2000 nonprofit groups working on watershed restoration in the United States from 1996 to 2008. The presence, number, and relative wealth of watershed groups was associated with improved dissolved oxygen in the water, as well as increased fishing and swimming amenities. Nonprofit groups tend to be local and often highly knowledgeable, and their presence also influences other watershed stakeholders to comply with governmental regulations. This analysis points to nonprofit organizations as an effective mechanism for helping to resolve large-scale collective action problems. —CA

Proc. Natl. Acad. Sci. U.S.A. 10.1073/pnas.1805336115 (2018).

ANTHROPOLOGY

Moral judgment across cultures

The importance of intentions in moral judgments varies across cultures. One theory holds that these differences are driven by variation in cultural norms regarding the acceptability of inferring other peoples' mental states. To test this question, McNamara *et al*. investigated moral judgments among indigenous iTaukei Fijians whose cultural norms discourage mental-state reasoning. Participants judged cases in which an actor transgressed by accident more harshly than cases in which an actor intended to do harm but failed to do so. When participants were primed to focus on thoughts in a follow-up study, this pattern reversed. These data suggest that moral and legal theories that emphasize intention may emerge from an implicit focus on mental states that is often taken for granted in Western contexts. —TSR

Cognition **182**, 95 (2019).

DIET

Temperature and food preference

On a cold day, hot soup seems nice. This desire has a fundamental evolutionary basis that may lie in survival. Brankatschk *et al*. provided *Drosophila* fly larvae with a choice of food types: one containing only yeast and another consisting only of plant material. At moderate temperatures, yeast was preferred, but at cold temperatures, the larvae's preference shifted to plant material. When these larvae metamorphosed into adult flies, they showed increased cold resistance and improved survival in the winter, possibly because the plant fatty acids enhance membrane fluidity and motor coordination. —BAP

Dev. Cell. **46**, 781 (2018).

STEM CELLS

Blood by the clock

Hematopoiesis is the process by which the cellular components of blood are maintained by hematopoietic stem and progenitor cells. Golan *et al*. found that supporting stem cells in the bone marrow and replenishing mature cells in the blood is coupled to the light-dark cycle. Light, via norepinephrine and tumor necrosis factor (TNF), induces blood replenishment and suppresses stem cell self-renewal, whereas darkness, via melatonin and TNF, promotes stem cell self-renewal and diminishes vascular permeability to increase the pool of cells available for the next day. Understanding the daily rhythm of the bone marrow could have implications for the time of day at which bone marrow and stem cell transplants are likely to be most successful. —GKA

Cell Stem Cell **23**, 572 (2018).

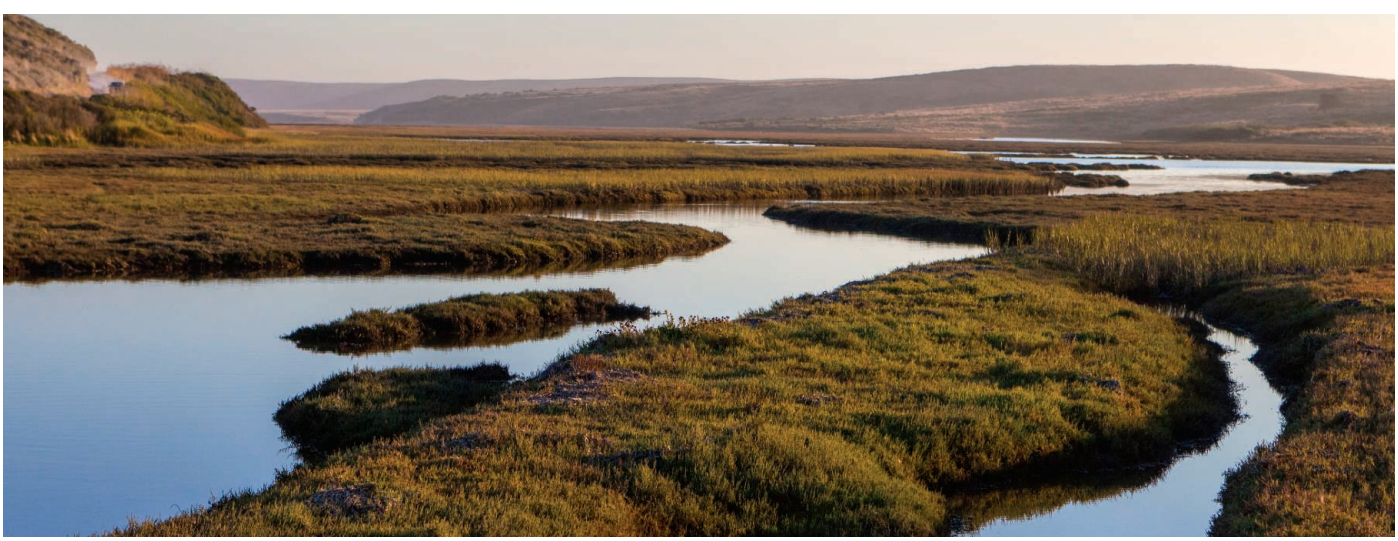


PHOTO: ULLSTEIN BILD/GETTY IMAGES

U.S. watershed restoration groups effectively convert donations into environmental cleanup.

ALSO IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

MAGNETIC MATERIALS

Faster switching for soft magnets

The most familiar magnets are permanent magnets like the ones on a refrigerator door. However, for applications in transformers and motors, soft magnets that can rapidly switch their magnetization in response to a magnetic field are used. In electronics, wide bandgap semiconductors such as silicon carbide will allow power conversion electronics and motor controllers to operate more efficiently, but soft magnets must be developed that can respond at higher frequencies. Silveyra *et al*. review the development of current soft magnetic materials and opportunities for improving their performance in high-frequency operation. Materials being explored include soft ferrites, amorphous and nanocrystalline alloys, and powder cores or soft magnetic composites. —PDS

Science, this issue p. 418

CHROMATIN STRUCTURE

Imaging chromatin spatial organization

The genome is organized within the nucleus as three-dimensional domains that modulate DNA-templated processes. Bintu *et al*. used high-throughput Oligopaint labeling and imaging to observe chromatin dynamics inside the nuclei of several different mammalian cell lines. After combining the datasets, single-cell matrices revealed chromatin arranged in topologically associating domains (TADs). Removing cohesin resulted in a loss of aggregate TADs among populations of cells, but specific TADs were still detected at the single-cell level. Furthermore, higher-order organization was detected, suggestive of cooperative interactions within the genome. —LMZ

Science, this issue p. 419

CANCER

Cancer chromatin accessibility landscape

The Cancer Genome Atlas (TCGA) provides a high-quality resource of molecular data on a large variety of human cancers. Corces *et al*. used a recently modified assay to profile chromatin accessibility to determine the accessible chromatin landscape in 410 TCGA samples from 23 cancer types (see the Perspective by Taipale). When the data were integrated with other omics data available for the same tumor samples, inherited risk loci for cancer predisposition were revealed, transcription factors and enhancers driving molecular subtypes of cancer with patient survival differences were identified, and noncoding mutations associated with clinical prognosis were discovered. —SYM

Science, this issue p. 420; see also p. 401

LIMB REGENERATION

How the axolotl makes a new limb

Unlike most vertebrate limbs, the axolotl limb regenerates the skeleton after amputation. Dermal and interstitial fibroblasts have been thought to provide sources for skeletal regeneration, but it has been unclear whether preexisting stem cells or dedifferentiation of fibroblasts formed the blastema. Gerber *et al*. developed transgenic reporter animals to compare periskeletal cell and fibroblast contributions to regeneration. Callus-forming periskeletal cells extended existing bone, but fibroblasts built new limb segments. Single-cell transcriptomics and Brainbow-based lineage tracing revealed the lack of a preexisting stem cell. Instead, the heterogeneous population of fibroblasts lost their adult features to form a multipotent skeletal progenitor expressing the embryonic limb program. —BAP

Science, this issue p. 421

STRUCTURAL BIOLOGY

Structure of the largest, most complex ribosome

Ribosomes are two-subunit ribonucleoprotein assemblies that catalyze the translation of messenger RNA into protein. Ribosomal RNAs (rRNAs) play key structural and functional roles. Ramrath *et al*. report the high-resolution structure of mitochondrial ribosomes from the unicellular parasite *Trypanosoma brucei* that contain the smallest known rRNAs. The trypanosomal mitoribosome is the most complex ribosomal assembly characterized, with two rRNAs and 126 proteins. The increased protein subunits have substituted for rRNA as an architectural scaffold. The structure also reveals the minimal core needed for ribosome function. —SYM

Science, this issue p. 422

ORGANIC CHEMISTRY

Salt and iron ease the way to amines

Adding nitrogen to carbon compounds is often a laborious process. The nitrogen typically needs protection to prevent undesired reactivity, and the protecting groups can be hard to remove afterward. Legnani *et al*. report a versatile method to add unadorned NH_2 from a hydroxylamine derivative directly to a double-bonded carbon center. A simple iron catalyst manipulates chloride from table salt onto the adjacent carbon through a radical mechanism, where it is poised for a wide variety of further functionalization options. —JSY

Science, this issue p. 434

GAS SEPARATION

A preference for ethane

Industrial production of ethylene requires its separation from ethane in a cryogenic process that consumes large amounts of energy. An alternative would be differential sorption in microporous materials. Most of these materials bind ethylene more strongly that ethane, but adsorption of ethane would be more efficient. Li *et al*. found that a metal-organic framework containing iron-peroxo sites bound ethane more strongly than ethylene and could be used to separate the gases at ambient conditions. —PDS

Science, this issue p. 443

QUANTUM OPTICS

Single photons from inflated atoms

Single-photon emitters are of interest for many applications, such as quantum sensing and quantum secure communication. Although efficient on-demand solid-state sources based on quantum dots, defect color centers in diamond, and single molecules are available, most of these systems require cryogenic temperatures for optimal performance. Ripka *et al*. used a cloud of warm excited rubidium atoms confined to a gas cell and exploited the exaggerated interaction of Rydberg states to generate single photons on demand. The results demonstrate the viability of atomic gases for application in the development of quantum technologies. —ISO

Science, this issue p. 446

MICROBIOME

Deconstructing probiotics

Besides supporting host metabolism, our intestinal microbiota also plays a vital role in modulating functions of immune cells in the gut. Verma *et al*. examined how a particular probiotic strain, *Bifidobacterium bifidum*, promotes the generation of regulatory T cells (T_{regs}) in the intestine. β-glucan/galactan polysaccharides derived from the cell wall of *B. bifidum* were responsible for promoting T_{reg} induction in the intestine. This process was dependent on intestinal dendritic cells that express Toll-like receptor 2. Thus, microbial components rather than

live microbes could potentially be used to treat microbial dysbiosis associated with gastrointestinal disorders, including colitis and Crohn's disease. —AB

Sci. Immunol. **3**, eaat6975 (2018).

CALCIUM SIGNALING

When organelles signal better together

Wolfram syndrome is caused by loss-of-function mutations in the endoplasmic reticulum (ER) protein WFS1. However, patients show symptoms typically associated with mitochondrial dysfunction, such as diabetes and optic atrophy. Angebault *et al*. found that patient fibroblasts had reduced mitochondria-ER contact sites, decreased mitochondrial calcium (Ca^{2+}) uptake, and lower mitochondrial respiration than control fibroblasts. WFS1 deficiency destabilized the Ca^{2+} sensor NCS1, and reconstituting NCS1 in patient fibroblasts restored mitochondrial respiration and Ca^{2+} signaling. —WW

Sci. Signal. **11**, eaaq1380 (2018).

EVOLUTION

Showing their stripes

The adaptive radiation of East African cichlids has led to more than 1200 species across a number of lakes. Across these species, many convergent traits have emerged, including the presence or absence of horizontal stripes. Kratochwil *et al*. show that the appearance or loss of stripes is related to changes in the *agouti-related peptide 2* gene, which acts as a kind of on-off switch for stripe generation (see the Perspective by Gante). This action has enabled rapid and repeated evolution of stripes across this speciose radiation. —SNV

Science, this issue p. 457; see also p. 396

REVIEW SUMMARY

MAGNETIC MATERIALS

Soft magnetic materials for a sustainable and electrified world

Josefina M. Silveyra, Enzo Ferrara, Dale L. Huber, Todd C. Monson*

BACKGROUND: Soft magnetic materials and their related devices (inductors, transformers, and electrical machines) are often overlooked; however, they play a key role in the conversion of energy throughout our world. Conversion of electrical power includes the bidirectional flow of energy between sources, storage, and the electrical grid and is accomplished via the use of power electronics. Electrical machines (motors and generators) transform mechanical energy into electrical energy and vice versa. The introduction of wide bandgap (WBG) semiconductors is allowing power conversion electronics and motor controllers to operate at much higher frequencies. This reduces the size requirements for passive components (inductors and capacitors) in power electronics and enables more-efficient, high–rotational speed electrical machines. However, none of the soft magnetic materials available today are able to unlock the full potential of WBG-based devices.

Since the 1800s, when iron was the only soft magnetic material available, metallurgists, materials scientists, and others have been periodically introducing improved materials. The invention of silicon (electrical) steel in 1900 was a notable event for soft magnetic materials. Silicon steel still dominates the global market of soft magnets and is the material of choice for large-scale transformers and electrical machines. However, its low electrical resistance makes it subject to large losses from eddy currents, particularly as operating frequencies are increased. This leaves the soft magnetic community looking to other materials to meet the needs of newer high-frequency devices.

ADVANCES: Several soft magnetic materials show promise for high-frequency operation. As oxides, soft ferrites stand out from other magnetic materials because they are insulating and therefore excel at reducing losses from eddy currents. However, soft ferrites suffer from a saturation magnetization (M_s) that is approximately one-fourth that of silicon steel. This substantially limits the power density in devices designed using soft ferrites and therefore limits their application. The current state-of-the-art materials are the amorphous and nanocrystalline alloys, which were invented in 1967 and 1988, respectively. Their distinctive nanostructures and extremely thin laminations work together to suppress eddy current losses, even at high frequencies. However, fabricating parts by cutting and then stacking or winding extremely thin and brittle laminations can be challenging. Composites are the newest class of soft magnetic materials, and in the field of soft magnetics, they are referred to as powder cores or soft magnetic composites. These materials are composed of micrometer-sized particles coated with an insulating binder and consolidated at high pressures. Although their performance is rather modest, their isotropic nature makes them well suited for use in rotating electrical machines.

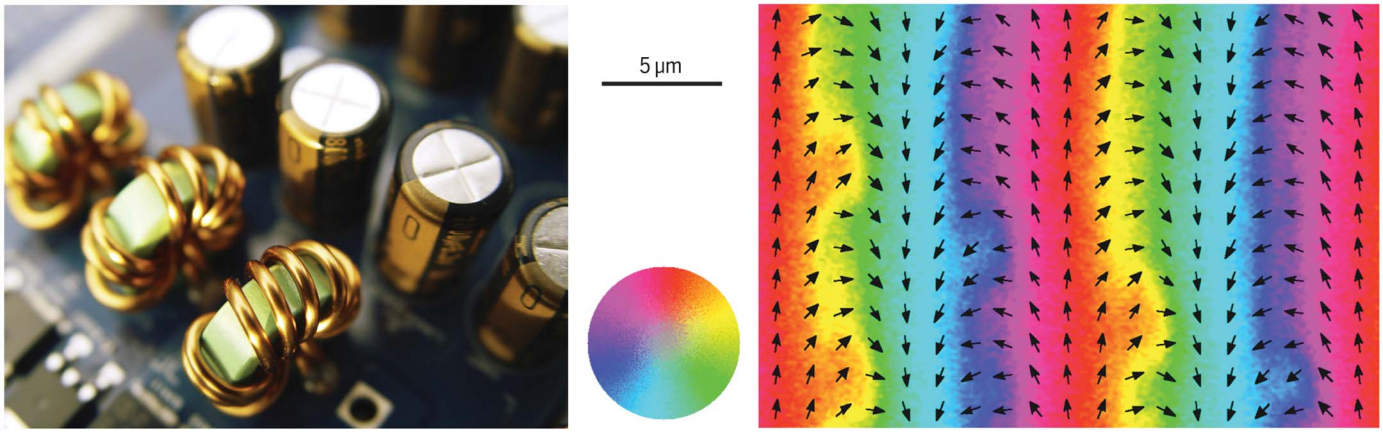
OUTLOOK: The need for improved soft magnets capable of efficient operation at high frequencies is capturing the attention of a growing number of researchers. Improvements to existing materials are being developed by some, whereas others are exploring radically new approaches. Even though ferrites were invented in the 1940s, grain-boundary engineering and new syntheses compatible with integrated manufacturing are being pursued. The nanocrystalline and amorphous materials are constantly being improved through increases in M_s and the introduction of alloys that are more amenable to the fabrication of large-scale parts. Powder cores have opened the door for nanoparticle-based composites, which can be fabricated with top-down as well as bottom-up approaches. A well-designed nanocomposite has the potential for negligible losses over specific temperature and frequency ranges. In all materials, advanced characterization techniques will be paramount to understanding the relationship between nanostructure and magnetization reversal, which will then enable the design of improved soft magnets. ■

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aao0195

The list of author affiliations is available in the full article online.
*Corresponding author. Email: tmonson@sandia.gov
Cite this article as J. M. Silveyra *et al.*, *Science* **362**, eaao0195 (2018). DOI: 10.1126/science.aao0195

Two views of soft magnetic materials. (Left) A row of three wound toroidal inductors mounted on a printed circuit board. (Right) Magneto-optical Kerr effect image of the magnetization pattern on the surface of an amorphous alloy soft magnet.



CREDITS: (LEFT TO RIGHT) BILLNOLL/ISTOCK; R. SCHAEFER *ET AL.*, *IEEE TRANS. MAGN.* **37**, 2245 (2001)

REVIEW

MAGNETIC MATERIALS

Soft magnetic materials for a sustainable and electrified world

Josefina M. Silveyra[1], Enzo Ferrara[2], Dale L. Huber[3], Todd C. Monson[4]*

Soft magnetic materials are key to the efficient operation of the next generation of power electronics and electrical machines (motors and generators). Many new materials have been introduced since Michael Faraday's discovery of magnetic induction, when iron was the only option. However, as wide bandgap semiconductor devices become more common in both power electronics and motor controllers, there is an urgent need to further improve soft magnetic materials. These improvements will be necessary to realize the full potential in efficiency, size, weight, and power of high-frequency power electronics and high–rotational speed electrical machines. Here we provide an introduction to the field of soft magnetic materials and their implementation in power electronics and electrical machines. Additionally, we review the most promising choices available today and describe emerging approaches to create even better soft magnetic materials.

As we transition to both a more electrified world and sustainable sources of energy, efficient power conversion will take on an increasingly important role to manage the bidirectional flow of power between sources (solar arrays, wind turbines, or other power plants), storage, and the electrical grid. Power conversion can also describe the transformation between mechanical energy and electrical energy, which is being accomplished to a greater extent with electrical machines (both motors and generators). Electrical machines are a vital part of industry worldwide and are becoming very important to all forms of transportation. In both cases of power conversion described above, the advancing field of wide bandgap (WBG) semiconductors is an enabling technology. Their high switching speed and increased operating temperature are driving faster and more compact power conversion electronics and motor controls. Additionally, WGB devices have decreased cooling requirements compared with silicon electronics.

In the field of power electronics, however, efforts to modernize the passive components (transformers, inductors, and capacitors) have lagged compared with the abundance of work focusing on semiconductor switches. Here we focus on the magnetic-based passive devices and address the challenges facing transformers and inductors, which both require a soft (low coercivity) magnetic core to achieve high power densities. Advances in the magnetic materials used in transformers and inductors will be equally beneficial to electrical machines. Thus, we will also discuss how electrical machines could realize substantial improvements in power density and efficiency through improved soft magnetic materials. Given all of the applications of power electronics and electrical machines, we will see that soft magnets are ubiquitous and that research focusing on improving soft magnetic materials could have a substantial impact on global energy efficiency.

Fundamentals of soft magnets

Before continuing our discussion of soft magnets, it is important to describe the basic characteristics of a magnetization curve and define some key magnetic parameters. The defining relation between the three magnetic field vectors is

$$\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M}) = \mu_0\mathbf{H} + \mathbf{J} \qquad (1)$$

where μ_0 is the magnetic permeability of vacuum, $\mathbf{B}$ is the magnetic flux density or induction, $\mathbf{H}$ is the magnetic field, $\mathbf{M}$ is the magnetization or volume density of the magnetic moment, and $\mathbf{J}$ ($\mu_0\mathbf{M}$) is the magnetic polarization. In materials science, it is more common to represent magnetic response in terms of $\mathbf{M}$, whereas device technology most often uses $\mathbf{J}$, which is analogous to the polarization ($\mathbf{P}$) in dielectric materials. The response of magnetic materials can be represented in terms of $\mathbf{M}(\mathbf{H})$, $\mathbf{B}(\mathbf{H})$, or $\mathbf{J}(\mathbf{H})$. Within this magnetic response is embedded a complex mix of processes that includes domain wall motion, rotation of magnetic moments, and eddy current losses (*1*).

Several key parameters for a soft magnetic material can be extracted from its magnetization curve. The magnetic susceptibility is defined as $\chi = M/H$ and the relative magnetic permeability is defined as $\mu_r = \mu/\mu_0 = B/\mu_0 H = (1 + \chi)$. The susceptibility (and also μ_r) can be measured at any point along the magnetization curve. However, the value of χ or μ_r is typically measured starting at the point of zero field and reported as the initial susceptibility or permeability (χ_i or μ_i, respectively). Both μ_r and χ are indications of how much magnetic field is required to change the direction of the magnetization in a sample. Other key parameters include a sample's saturation magnetization (M_s), which is the sample's maximum magnetization, magnetic remanence (M_r), or magnetization that remains after all external field is removed, and coercive field (H_c), or the strength of the opposing field required to remove a sample's magnetization after saturation. Figure 1 displays a magnetization curve, $M(H)$, for a soft magnetic material, with M_s, M_r, H_c, and χ_i labeled.

Three other properties of magnetic materials are also important descriptors. One such property is a magnetic material's anisotropy (K), which is the energy difference between the preferred and nonpreferred directions of magnetization within a material. Magnetocrystalline anisotropy is an intrinsic property of a material determined by its crystalline structure and spin-orbit coupling. Although magnetocrystalline anisotropy can be described with series expansions with terms out to many orders, most often the first-order constant (K_1) is sufficient for calculations. The second property is magnetostriction (λ), which is a material's dimensional change as a function of applied magnetic field. Magnetostriction can be a source of loss through the unwanted conversion of electromagnetic energy into mechanical energy. This effect results in the humming sound created by the magnetic cores of transformers. Additionally, there can be a coupling between a material's magnetization and strain (magnetoelastic coupling). This coupling can result in an additional anisotropy term, stress anisotropy (K_σ). A material's overall or effective anisotropy (K_{eff}) approximately follows a proportional relation with H_c, whereas the initial susceptibility χ_i is inversely proportional to K (*2*).

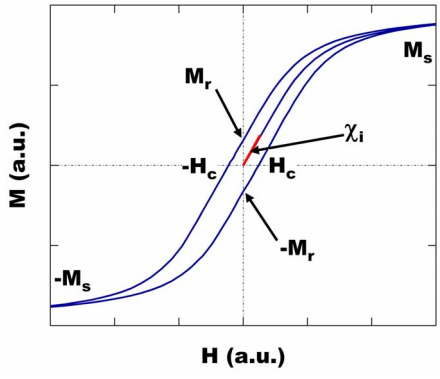


Fig. 1. Magnetization curve for a soft magnet. A magnetization curve, $M(H)$, is depicted for a soft magnetic material with M_s, M_r, H_c, and χ_i labeled. Arbitrary units (a.u.) are used for both M and H, as the intention of this plot is to clearly illustrate the different key magnetic parameters.

[1]Laboratorio de Sólidos Amorfos, INTECIN, Facultad de Ingeniería, Universidad de Buenos Aires - CONICET, Buenos Aires, Argentina. [2]Istituto Nazionale di Ricerca Metrologica (INRIM), Torino, Italy. [3]Center for Integrated Nanotechnologies, Sandia National Laboratories, Albuquerque, NM 87185, USA. [4]Sandia National Laboratories, Albuquerque, NM 87185, USA.
*Corresponding author. Email: tmonson@sandia.gov

Finally, the Curie temperature (T_c) is the temperature above which a material loses its magnetic order and becomes paramagnetic. T_c is an important parameter for power electronics and electrical machines, which often operate at high temperatures and where space for active cooling may not be available.

When people consider magnets, they typically think about permanent (hard) magnets. Permanent magnets have a large H_c (>1000 A/m) and high M_r that allows them to retain a large magnetization after being magnetized. This property allows hard magnets to attract and repel other magnets, as well as to magnetize other materials (such as the steel in your refrigerator door). However, for the inductive applications found in power electronics and electrical machines, permanent magnets are not useful. Instead, these applications require soft magnets that have a low H_c (<1000 A/m) and very little M_r. Inductive applications are based on the ability to rapidly switch the magnetization of a material with a magnetic field created by current-carrying coils of wire. The high μ_r of a soft magnet concentrates (by orders of magnitude greater than that of an air core) the magnetic field lines inside the windings of an inductor or electrical machine and boosts the performance of the inductive device by allowing it to store more energy in the form of magnetic flux density. An increase in energy density is precisely what is needed for emerging power electronics and electrical machines for which size, weight, and power (SWaP) are often the most important metrics. For this reason, increasing the M_s of soft magnets can have a positive impact on SWaP by increasing the maximum field at which a magnetic device can operate.

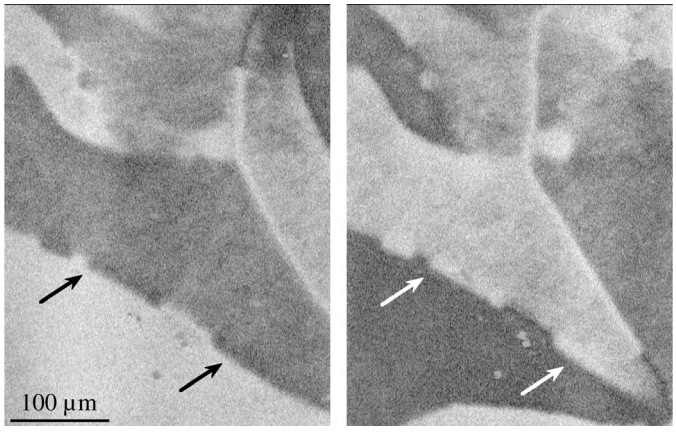


Fig. 2. Pinned domain wall. A pinned domain wall in an amorphous alloy ribbon is marked by arrows in both the left and right Kerr microscopy images. The position of the pinned wall remains fixed, whereas the surrounding domain structure changes from left to right. [Reproduced from (*94*)]

Energy losses

The increase in energy density that soft magnetic core materials bring to inductive devices does not come without a cost, as core materials can also be a place of energy loss, particularly as operating frequencies increase. Hysteretic losses occur through a magnetic material's coercivity. Every time a magnet completes a full cycle of its magnetization curve, the area inside the curve is a measure of the energy lost. Thus, as operatng frequencies increase, even magnets with low coercivity can become quite lossy. The second major loss mechanism for soft magnetic devices is eddy currents, which are closed paths of electrical current generated in a conductor by a time-varying magnetic field. These current loops create a magnetic field in opposition to the change in magnetic flux (according to Faraday's law of induction). Power losses from eddy currents roughly scale as the operating frequency squared, so they can become the dominant source of energy loss at high switching rates. Mitigating both hysteretic and eddy current losses consumes a large portion of the resources dedicated toward improving soft magnetic materials.

Theoretical treatment of magnetic losses can be quite complicated, as magnetic materials are not continuous and instead contain magnetic domains, domain walls, and a complex distribution of eddy currents. One of the first models on magnetic losses was developed by Williams, Shockley, and Kittel (*3*). Their model, which only applied to 180° Bloch walls in monocrystalline silicon steel, was broadened to describe loss behavior in a grain-oriented silicon steel sheet by Pry and Bean (*4*). Bertotti took a statistical approach and applied the concept of loss separation in developing the statistical theory of losses (STL) (*5*). According to the STL, energy loss, $W(f)$, at a given magnetizing frequency f and peak magnetic polarization J_p, is the sum of three components: hysteresis loss (W_h), classical loss (W_{cl}), and excess losses (W_{exc}). These terms correspond, respectively, to the residual energy dissipated in the limit $f \to 0$; the loss component calculated by Maxwell's equations, assuming the material is magnetically homogeneous (absent of domains); and the energy loss associated with the large-scale motion of the domain walls (*5*)

$$W(f) = W_h + W_{cl}(f) + W_{exc}(f) \quad (2)$$

Although all three components, at their root cause, are the result of eddy current mechanisms (only at different space-time scales) (*1*), the classical loss is the term most closely associated with eddy current losses and can be written as

$$W_{cl} = (\pi^2/6) \cdot \sigma d^2 J_p f \quad (3)$$

where σ is the conductivity of the material and d is the lamination thickness (assuming that d is greater than the skin depth, which may not be the case at high frequencies). The squared dependence on d is the reason many soft magnetic materials—in particular, steels, amorphous alloys, and nanocrystalline alloys—are fabricated into parts as wound tape or stacked laminations. Excess loss, which is associated with large-scale domain wall motion, is described in more detail elsewhere but increases as a function of $f^{1/2}$ (*6*–*8*). Because a substantial portion of losses arises from domain wall motion, developing finer control over the micro- and nanostructure of magnetic materials can reduce impediments to domain wall motion and overall losses. Figure 2 demonstrates an example of an impediment to domain wall motion, capturing two different pinning sites via the use of magneto-optical Kerr effect (MOKE) imaging.

The STL complies, in general, with experiments up to frequencies at which magnetic shielding strongly enters into play. Beyond this limit (>1 MHz), more complex formulations, exploiting hysteresis modeling (such as the dynamic Preisach model), and numerical calculations are usually employed (*9*–*12*). Measurements can consistently proceed up to radio frequencies, starting from quasi-static characterization, and the theoretical approach can be fully applied under a quasi-linear constitutive magnetic equation, which implies low or very low J_p (within the Rayleigh region).

Power electronics

Power electronics enable the efficient conversion, control, and conditioning of electric energy between the source and a load (*13*). The benefits of power electronics include improved power quality, voltage regulation, improved response to fault conditions, increased operational flexibility, and reduced interconnection costs (*14*). The improved efficiency of modern power electronics, if implemented worldwide, has the potential to reduce global energy consumption by as much as 20% (*15*). Power electronics can be found almost everywhere: Buildings and lighting, the electrical grid, drives for electric motors and generators, and power supplies for consumer and industrial appliances are just a few of the most likely applications. The invention of the mercury-arc valve in 1901 was the beginning of power electronics, and with the invention of silicon-based diodes in the late 1940s the field entered the semiconductor age (*16*). Power electronics then became much more widespread after the invention of silicon-based metal oxide semiconductor field-effect transistors and insulated-gate bipolar transistors.

Increased demand for efficiency and the need to withstand harsher operating environments led to the investigation of semiconductor materials beyond silicon—in particular, the WBG materials SiC and GaN. Stronger atomic bonds

in these compounds result in wider energy gaps (3.2 to 3.4 eV versus 1.1 eV for Si) and breakdown fields 10 times as high as those of silicon (*17*). Thus, WBG devices have "on" resistances and related losses that are 1/10 the amount in Si-based devices (*18*). These properties also allow smaller device areas, which reduces both the capacitance and the resistor-capacitor time constant. The result is lower losses per switching cycle, which in turn enables operation at higher frequencies. Additionally, WBG devices can operate at temperatures well above 300°C, twice as high as the maximum operating temperature for silicon (*17*).

A key benefit of the higher operating frequencies enabled by WBG semiconductors is a decrease in the inductive and capacitive requirements for a given circuit. As an example, the undesired alternating voltage component, or ripple voltage (V_{ripple}), in a step-down (buck) dc-to-dc converter is given by

$$\frac{V_{ripple}}{V_{out}} = \frac{1-D}{8LCf^2} \quad (4)$$

where V_{out} is the converter output voltage, D is the duty cycle, L is the circuit inductance, and C is the capacitance. The size of the inductor and capacitor can be reduced substantially while keeping the ripple voltage constant by increasing the operating frequency. Similar reductions in the footprint of passive devices can be realized in almost all power electronics circuitry.

Advances in power electronics are also enabling the development of intelligent and agile solid-state transformers, which operate at a much higher frequency (>1 kHz) and shrink the size of the conventional 60-Hz transformer by a factor of 10 or more (*19*, *20*). Although WBG semiconductors enable the reduction of soft magnetic components, none of the magnetic materials available today can fully unlock the potential of WBG devices (*21*).

Electrical machines

Electric motor–driven systems are the single largest consumer of electricity, accounting for more than 40% of the electrical energy generated globally (*22*). Even small efficiency improvements in electric machines will have a large impact on energy savings. According to the International Energy Agency, if by 2030 all industries adopted the already available high-efficiency technologies for their electric motor–driven systems, energy savings would exceed 300 billion kWh, thus helping to reduce CO_2 emissions by 200 billion kg per year (*22*). There would also be financial benefits from more efficient motors, because energy consumption is typically responsible for more than 90% of the overall cost of a motor (*22*, *23*). Electric cars, aircraft, and spacecraft need higher efficiencies and power densities because energy storage is limited and volume and weight are also crucial.

As electric machines get smaller, design engineers are facing the growing challenge of keeping the components cool. Overheating deteriorates the properties of most materials (such as insulators, coils, and permanent magnets) and reduces the service life of the device. Fewer losses would reduce cooling loads, permit the use of a smaller fan at the end of the rotor, and further decrease the size of the machine.

In the conversion between electrical energy and mechanical energy, electrical machines experience several loss mechanisms, including mechanical, winding or coil (also referred to as copper), and core (also referred to as iron). Core losses, which are due to ferromagnetic hysteresis and eddy currents in the stator and rotor, are independent of load. They depend on the properties of the magnetic material, the flux density, and the rate of change of the flux density (i.e., switching frequency). Core losses have typically been estimated to account for 15 to 25% of the total loss (*24*). However, this percentage increases substantially as machines are operated at higher speeds. For example, core losses account for 59% of the total loss in a 110-kW machine operating at 51,000 revolutions per minute (rpm) (*25*).

Operation at much higher speeds, which is enabled by advanced motor controls driven by new WBG semiconductors, is another approach to increasing power. The mechanical power of a motor is defined as the rotational speed multiplied by the torque, or $\omega \cdot T$. By increasing rotational speed, either the power density can be increased or the size of the motor can be decreased, which leads to a reduction of the critical rare-earth elements in permanent magnet motors (*26*). Additionally, gearboxes can be eliminated in various applications that require high rotational speeds. A major problem with using high speeds is that core losses increase as a function of the switching frequency, which is where advances in soft magnetic materials can influence performance improvements in electrical machines.

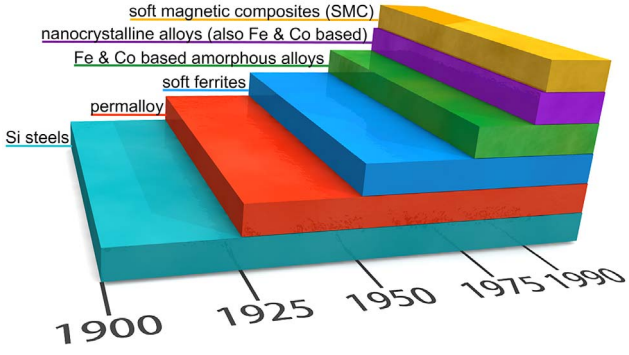


Fig. 3. A brief history of soft magnetic materials. A synopsis of the progression of soft magnetic materials is shown. This chart is by no means exhaustive. [Modified from (*96*), with permission]

Brief history of soft magnetic materials

Since Michael Faraday demonstrated electromagnetic induction in 1831, soft magnetic materials have continued to evolve. Faraday's natural choice of core material was iron, which has the highest room temperature M_s of any element, in addition to a large μ_r and fairly low H_c. However, there was room for considerable improvement, even in a simple material composed of a single element. It was discovered that annealing iron not only improved its mechanical properties but also decreased its coercivity through stress relief, making it better suited for use in inductive applications.

Seeking even better performance, scientists and engineers looked for ways to improve upon the properties of soft iron. In 1900, Robert Hadfield, a metallurgist from England, invented nonoriented silicon steel by adding up to 3% silicon to iron and increasing its electrical resistivity (ρ) while also increasing μ_r (*27*). American metallurgist Norman Goss invented grain-oriented silicon steel in 1933 by promoting grain growth along a crystalline direction of low anisotropy, increasing μ_r even further (*27*). Even today, silicon (or electrical) steels account for a major share of the global soft-magnet market because of their high M_s and relatively low cost (*1*). The most common applications for silicon steel are large-scale transformers (grain-oriented silicon steel) and electrical machines (isotropic nonoriented silicon steel is preferred for rotating machines), for which its economical price is a huge benefit. However, a low ρ (~0.5 microohm·m) (*28*) makes silicon steels lossy at high frequency. Recently, electrical steel manufacturers have developed a path to increase the silicon content of their steel to 6.5% by using a chemical vapor deposition process (*29*). This approach increases ρ to 0.82 microohm·m but still leaves other materials as better choices for high-frequency power electronics and high–rotational speed electrical machines.

In the 1910s, Gustav Elmen at Bell Laboratories experimented with nickel-iron alloys and discovered the nickel-rich (78%) permalloy composition (*30*). A major advantage of permalloy is its high μ_r (up to 100,000). Nickel-iron alloys are still used in some specialty inductive applications today but are not common in power electronics and electrical machines because they have high eddy current losses, and the addition of nickel decreases M_s. With the addition of a small amount of molybdenum (2%) to permalloy, molypermalloy powder (MPP) can be produced (*31*). MPP is used to fabricate the lowest-loss powder cores (*31*, *32*), which will be discussed in more detail below.

In the late 1940s, magnetically soft ferrites were invented by J. L. Snoek (*33*). These materials are competitive because of their very high electrical resistivities (10 to 10^8 microohm·m), which make them effective at suppressing eddy current

losses. Additionally, because they are produced with ceramic processing techniques and abundant materials, ferrite parts can be produced at a very low cost. The high ρ and affordability of soft ferrites keeps these materials in high demand for inductive applications, including those at high frequency. In fact, their share of the global market in soft magnets is second only to silicon steel (*1*). They do suffer from a relatively low M_s (nearly a quarter of that of silicon steel), which limits the energy density of inductive elements containing a ferrite core.

In 1967, a new class of materials, amorphous alloys, was invented (*34*). By the mid-1970s, interest in iron- and cobalt-based amorphous alloys was surging, and these materials began finding their way into applications. Through the elimination of any long-range order, coercivity is substantially reduced in these alloys. In 1988, researchers at Hitachi included Nb and Cu additives and added an annealing step to the production of amorphous alloys to produce small and closely spaced crystallites of iron or cobalt (on the order of 10 nm in diameter) within a matrix of amorphous material (*35*). This was the inception of the nanocrystalline soft magnetic alloys. The formation of isolated transition metal crystallites reduced the eddy current losses of these materials in comparison with amorphous alloys. Both amorphous and nanocrystalline alloys are gaining market share in high-frequency power electronics and electrical machines today because of their low losses and competitive M_s. Despite a higher initial cost than silicon steel, these advanced alloys can reduce the total lifetime costs of power electronics and electrical machines, on account of reduced losses.

In the early 1990s, powder cores (also known as soft magnetic composites or SMCs) gained acceptance in some soft magnetic applications (*1*, *36*). These materials combine magnetic particles, anywhere between ~1 to 500 μm in diameter, and either coat or mix them with an insulating material before consolidating with high pressures (megapascal to even gigapascal pressures). Heat can also be applied either during or after densification to improve magnetic properties. The magnetic particles are most often Fe powders but can also consist of alloys such as MPP (mentioned earlier), Fe-P, Fe-Si, or Fe-Co. Because of the insulating and nonmagnetic matrix phase, these materials have a distributed air gap that limits their μ_r to a range of 100 to 500. However, the insulating matrix also boosts their ρ (10^{-3} to 10^{-1} microohm·m), reducing eddy current losses. SMCs can also be pressed into more complex final geometries without the need of any machining (net-shaping), which can substantially reduce manufacturing costs. Their isotropic nature, low cost, and ability to net-shape complex parts have made SMCs fairly successful in rotating electrical machines (*37*, *38*).

The brief history of soft magnetic materials described above (which is summarized graphically in Fig. 3) is by no means exhaustive. Instead, our intent is to focus on materials that have been and will continue to be competitive for the fabrication of soft magnetic components in high-frequency power electronics and electrical machines. Performance metrics such as M_s and core loss are extremely important. However, because soft magnetic parts will need to be used in large quantities, the importance of cost cannot be neglected. For this reason, soft ferrites still remain a competitive core material at high frequency. Because of their excellent performance at high frequency, the amorphous and nanocrystalline alloys will certainly continue to be key materials. Although silicon steels still make up a majority of the global market for soft magnetic materials, their primary applications are in large transformers operating at 50 or 60 Hz and slow–rotational speed electrical machines. Thus, we will not discuss silicon steel further. We will, however, consider the development of new soft magnetic materials, including both top-down and bottom-up approaches that could build on a foundation of work established in the area of SMCs.

Soft ferrites

Ferrites are distinctive soft magnetic materials in that they are ionic compounds. They are also ferrimagnetic, whereas all of the other soft magnetic materials discussed are ferromagnetic. J. L. Snoek and co-workers developed ferrites between 1933 and 1945 at the Philips Research Laboratories in Eindhoven, Netherlands, where commercial cores were required for Pupin coils in telephone cables and loudspeakers (*33*). They investigated the properties of magnetic oxides of the chemical formula $MO{\cdot}Fe_2O_3$, where M is a divalent ion (such as Fe^{2+}, Mn^{2+}, Ni^{2+}, Zn^{2+}, or Mg^{2+}) known to occupy two different kinds of crystallographic positions (A and B sites) within the oxide cell. L. Néel made the basic assumption that the exchange force acting between ions on the A and B sites was negative, providing the theoretical key to an understanding of the phenomena of ferrimagnetism and antiferromagnetism in ferrites, where magnetization is caused by the magnetic moments of the metal ions, but not according to the conventional ferromagnetic exchange interaction (*39*).

The commercial production of these metal oxides is typically achieved with ceramic processing, starting with the calcination of either iron salts, base oxides, or metal carbonates. The starting reagents are thoroughly mixed by prolonged wet grinding until a homogeneous fine powder is obtained with granule dimensions in the micrometer range. Additional calcining and grinding steps may be applied at this point. Finally, the powders are mixed with a binder, pressed at high pressures, and then sintered at temperatures up to 1400°C in a controlled atmosphere. The final product, hard and brittle, is a semiconductor with an electrical resistivity six orders of magnitude higher than in Fe-based metallic alloys. Thus, even when a rapidly changing magnetic field is applied, eddy currents and energy loss are contained. This property makes ferrites suitable for high-frequency applications, and they are currently the most extensively used magnetic material up to frequencies of a few hundred megahertz. A collection of toroidal ferrite cores is displayed in Fig. 4.

Figure 5 shows one-eighth of single unit of the cubic spinel (named after the mineral spinel, $MgO{\cdot}Al_2O_3$) ferrite structure: The metallic cations are separated by the oxygen anions (O^{2-}) arranged in a face-centered cubic (fcc) structure; the small metal ions occupy interstitial positions, at either tetrahedral (A) or octahedral (B) sites, surrounded by four or six oxygen ions, respectively. Direct exchange interaction between the 3d electron spins of the metallic cations is negligible, whereas an indirect coupling mechanism, superexchange, occurs. The superexchange interaction is stronger when a straight line connects the cations through the O^{2-} ion (*40*). The A-B coupling, which is associated with an A-O-B angle around 125°, is much stronger than the A-A and B-B couplings through oxygen, whose angles are 90° and 80°, respectively. It involves the spins of the two extra 2p electrons in O^{2-}, which interact with the 3d spins of the two neighboring

Fig. 4. Toroidal ferrite magnetic cores. Shown are insulated commercial Mn-Zn ferrites (blue toroids) and Ni-Zn ferrites (pink, white) along with uncoated ferrites (black) of both types. Two insulated Ni-Zn toroids and one uncoated Mn-Zn toroid are prepared with primary (inner) and secondary (outer) windings for fluxmetric measurements.

metal cations. The mediating effect of the oppositely directed oxygen spins is such that the A and B ions' total magnetic moments are bound, according to Hund's rule, to antiparallel directions.

Ferrites have an antiferromagnetic configuration of the two sublattices corresponding with A ions spontaneously magnetized in one direction and B ions magnetized in the opposite direction. When the magnitudes of A and B spins are not equal, the material becomes a ferrimagnet and a net spontaneous magnetization results. The distribution of the divalent ions on A and B sites (and thus the magnetic properties) can be further modified by heat treatment, although this may depend on whether the material is quenched from a high temperature or slowly cooled. The ferrimagnetic nature of ferrites results in a relatively low M_s, typically 0.4 MA/m or less at room temperature. In addition, M_s in ferrites is more temperature sensitive than in ferromagnetic materials (*1*, *2*, *27*).

Ferrites can contain two or more different divalent ions, producing a mixed ferrite. The most common commercial soft ferrites have mixed cations, (Ni, Zn)O·Fe_2O_3, and (Mn, Zn)O·Fe_2O_3, but suppliers do not typically share the exact composition of their ferrites and instead provide performance specifications. In an interesting interplay between the ions within the spinel structure, the addition of nonmagnetic (*41*) ZnO·Fe_2O_3 to mixed ferrites actually increases their M_s. The presence of two or more metal ions M^{2+} provides great versatility of the magnetic properties. By making suitable additions and thermal treatments, material tailoring to specific applications can be achieved. For example, the mixed ferrites display a cubic symmetry with a negative value of the anisotropy constant, K. By adjusting composition and processing, very low anisotropy (and therefore low H_c) can be obtained over a range of temperatures suitable for many applications (20 to 100°C). Additionally, the temperature stability of μ_r can also be controlled by the addition of M^{2+} ions, such as Fe^{2+} or Co^{2+}. The highest μ_r of ferrospinels are found in the Mn-Zn ferrites, whereas Ni-Zn ferrites have a much higher ρ, depending on the amount of doping with Fe^{2+} (*42*).

The near-insulating character of ferrites is conducive to a nearly constant value of μ_r over many frequency decades, typically up to the megahertz region in Mn-Zn ferrites (ρ = 10 to 10^4 microohm·m) and the 10-MHz region in Ni-Zn ferrites (ρ as high as 10^8 microohm·m). This is illustrated in Fig. 6A, where the behavior versus frequency of the real component (μ′) of the relative initial permeability is presented for a commercial Mn-Zn ferrite at a peak polarization, J_p, ranging between 2 and 300 mT. The rather sudden drop in μ′ is brought on by the onset of electron spin resonance and is matched by a prominent peak in the imaginary permeability (μ″). See Fig. 6B for a plot of μ″ for values of J_p ranging between 2 and 300 mT. This response cannot be observed in other soft magnets because of their much lower ρ.

For application in the microwave region (f = 0.5 to 100 GHz), larger magnetic anisotropies are required. In this high-f range, a technical limitation was stated by Snoek (*43*) in 1948. Snoek observed that the product of dc initial susceptibility, χ_i, and resonance frequency, f_0, is a constant value related to spin precession and, thus, the magnetocrystalline anisotropy, K. In a bulk polycrystalline material, Snoek's law is written as

$$\chi_i f_0 = (1/2\pi)\gamma\mu_0 M_s \qquad (5)$$

where γ is the gyromagnetic ratio. Experimentally, the critical frequency may be taken as that at which the imaginary susceptibility χ″ peaks (Fig. 6B).

Some ferrites, such as Ba M-type (BaM) ferrites, having the same hexagonal structure of magnetoplumbite $BaO(Fe_2O_6)_3$, consist of spinel blocks rotated 180° with respect to one another and separated by an atomic plane containing the Ba atoms. This plane breaks the crystal symmetry and results in a hexagonal structure with a high K, thus shifting Snoek's limit up to 36 GHz. The large K of BaM ferrites can be combined with growth on substrates that introduce a high degree of crystal texture, such as MgO layers deposited on single-crystalline SiC (*44*), allowing the use of hexagonal ferrites in a variety of microwave devices.

With their high ρ, constant μ_r that extends well into the megahertz frequency range, and low cost, ferrites will continue to play a role in applications for which only low values of J_p (and low energy density) are required. Although ferrites have been around for a long time, strategies to improve their performance, including grain-boundary engineering (*45*) and embedding them in substrate materials (*46*), are being considered.

Amorphous and nanocrystalline alloys

In 1967, Duwez and Lin reported the first amorphous soft magnetic alloy in the form of small disk-shaped samples. They used a rapid solidification technique called splat cooling for the Fe-P-C system (*34*). Alloy composition engineering and the development of planar flow casting led to the production of amorphous ribbons from 5 to 50 μm in thickness and up to 25.4 cm in width (Fig. 7). Rapid cooling rates of 10^6°C/s freeze the alloy into an amorphous structure. As reported by Yoshizawa *et al.* in 1988, optimized compositions allow partial devitrification of the metallic glass during a controlled heat treatment. This process results in a nanocrystalline soft magnetic material that consists of ~10-nm diameter ferromagnetic nanograins embedded

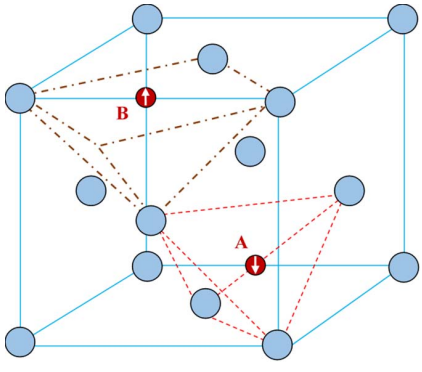


Fig. 5. One-eighth portion of the unit cell of a cubic spinel ferrite (ferrospinel). The O^{2-} ions (blue) are arranged in a fcc lattice. The metal ions (red) are arranged in either tetrahedral (A) or octahedral (B) interstitial sites, having ferrimagnetic (unbalanced antiferromagnetic) spin. The A-B coupling mediated by the oxygen ion leads to antiparallel configuration of metallic spins in tetrahedral and octahedral sites. [Modified from (*1*), with permission]

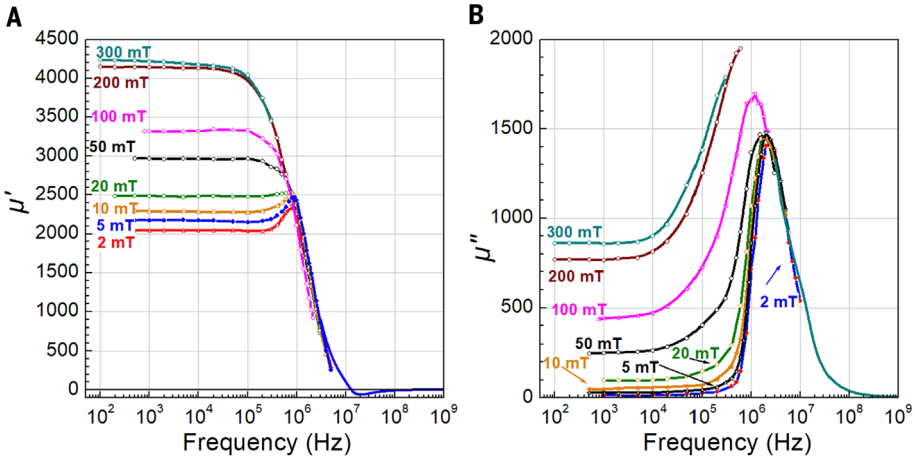


Fig. 6. Soft ferrite permeability response. (**A** and **B**) Measurements of (A) the real part (μ′) and (B) the imaginary part (μ″) of permeability versus frequency for a commercial (N87) Mn-Zn ferrite at peak magnetic polarizations, J_p, ranging from 2 to 300 mT. Permeability curves tend to coalesce above f ~ 1 MHz, where relaxation of the domain wall displacements is completed and only spin rotation processes survive. [Courtesy of F. Fiorillo, C. Beatrice, and S. Dobák]

in an amorphous matrix (i.e., glass ceramic) (*35*). Alben *et al*. explained that when the magnetic correlation length is longer than the structural correlation length, as in amorphous ferromagnetic materials, the effective magnetocrystalline anisotropy becomes negligible (*47*). This random anisotropy model was later extended to nanocrystalline materials by Herzer (*48*, *49*): Ferromagnetic nanocrystals are magnetically coupled through the ferromagnetic amorphous matrix, making the effective anisotropy very small (Fig. 8). Given the relatively high ρ of both amorphous and nanocrystalline alloys relative to crystalline materials and their very thin laminations, eddy current losses in these materials are very low, and they can operate safely into the tens-of-kilohertz frequency range (*50*). It should be noted that multiphase resistivity models are necessary to understand the compositional dependence on the ρ in nanocrystalline materials (*51*).

FINEMET was the first family of nanocrystalline alloys to be commercialized. Its most common composition is $Fe_{73.5}Si_{13.5}B_9Nb_3Cu_1$ (*35*), which is based on the amorphous Fe-Si-B system, where Si and B enhance glass formation. During the devitrification process (typically a 1-hour annealing at ~540°C) (*52*), Cu favors the nucleation of segregated Fe-Si crystals, whereas the large Nb atoms control grain growth and prevent the crystallization of anisotropic boride phases. The composite structure of FINEMET-type alloys has another strength: Fe-Si nanocrystals have a negative magnetostriction that balances the positive magnetostriction of the amorphous matrix, diminishing stress anisotropy (K_σ). The combination of vanishing magnetocrystalline and stress anisotropies results in a very low H_c (<10 A/m) and a very high initial μ_r (10^4 to 10^6). M_s ranges from ~0.9 to 1.0 MA/m. Yet instead of being a prejudicial factor, properly controlled magnetic anisotropies can be a powerful tool for tuning μ_r according to the application requirements. To this end, field annealing can be used to achieve uniform uniaxial anisotropies: Magnetic field annealing induces anisotropy by directional atomic ordering, whereas stress field annealing causes a plastic deformation of the material (*49*). Nevertheless, nanocrystalline Fe-based alloys are extremely brittle. Thus, they must be annealed in the final core geometry and must be handled carefully.

Because of their attractive characteristics, amorphous and nanocrystalline soft magnetic materials became a research focus in the years following their invention (*53*). The addition of Co to develop FeCo-based alloys (*54*) served two purposes. First, it increases M_s to 1.0 to 1.2 MA/m. Second, Co increases the Curie temperature of the amorphous and crystalline phases and allows higher operating temperatures. However, the increased magnetostriction of FeCo-based alloys makes them sensitive to manufacturing processes such as impregnation and cutting. When the iron is removed, pure Co-based alloys exhibit reduced M_s (0.6 to 0.8 MA/m) but are also suitable for high-temperature applications. Their composition can be tuned to exhibit near-zero magnetostriction or to enhance responses to anisotropic magnetic and stress fields during annealing, achieving μ_r values from 10 to 10^4 as needed for a given application (*55*). These cobalt-rich alloys can also have a markedly improved ductility, even in the nanocrystalline state. FeNi-based nanocrystalline alloys share many of the benefits of FeCo- and Co-based alloys: saturation inductions and Curie temperatures suitable for power applications, low losses at high switching frequencies, tunable permeability, and improved ductility (*56*). Yet because Ni is cheaper than Co, FeNi-based alloys have attracted recent efforts to explore their application in electric machines (*57*). Additionally, the modification of glass-former content in amorphous and nanocrystalline alloys has led to the optimization of other characteristics such as corrosion (*58*, *59*) and oxidation resistance (*60*), raw materials costs (*61*), and M_s [up to 1.5 MA/m (*62*, *63*)].

All families of amorphous and nanocrystalline materials have a common disadvantage: They are mechanically very hard (typically 800 to 1000 Vickers hardness), thus making machining a difficult and expensive task. If a conventional stamping process were used, dies would wear out quickly. These fabrication issues have held back their application in complex geometries for electric machines.

The first attempts to use amorphous materials for electric motors date back to 1981, just after the 1970s energy crisis. Mischler *et al.* demonstrated the low loss potential of the amorphous stator, but the fabrication technique required improvement (*64*). Since then, researchers in industry and academia have devoted their efforts to improving the fabrication of electric motors using amorphous and nanocrystalline soft magnetic materials. Given that either stacking or winding amorphous ribbons yield similar losses (*65*), researchers have applied both approaches in combination with cutting and machining. Additional methods used to cut mechanically hard amorphous sheets include chemical etching (*66*), wire electric discharge machining (EDM), and laser cutting (*67*). Laser cutting, however, has been shown to overheat and deteriorate the edges of Fe-rich amorphous materials (*68*). Grinding, EDM (*69*, *70*), or water-jet cutting (*71*) have also been used to machine already-assembled ribbons into the final core geometry, speeding up the manufacturing process.

When manufacturing challenges can be overcome, losses can be substantially reduced. Figure 9 shows an example in which finite-element analysis was used to calculate the distribution of losses in motor stators fabricated with silicon steel and Co-rich amorphous alloy. Axial motors can be particularly suited for amorphous laminations, as demonstrated by Hitachi with a 11-kW prototype motor that achieved an efficiency greater than that of the IE5 efficiency class for induction motors (*72*). Recently, some groups have also demonstrated the feasibility of constructing highly efficient electric motors with nanocrystalline cores (*73*, *74*).

Fig. 7. Ribbon of amorphous soft magnetic alloy. Thin (5 to 50 μm) ribbons of amorphous alloys are produced by planar flow casting.

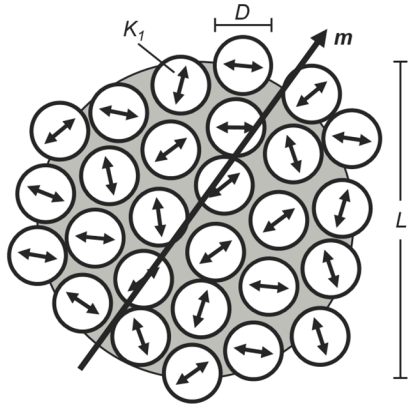


Fig. 8. Random anisotropy model for nanocrystalline materials. Double arrows indicate the randomly oriented anisotropy (K_1 is the first-order anisotropy constant) within the structural correlation length (grain size, *D*). The long arrow indicates the orientation of the magnetization (**m**) within the magnetic correlation length (*L*). [Reproduced from (*53*)]

Amorphous core transformers have been commercially available since the 1980s. A decade later, 6 to 8% of the distribution transformers bought in the United States were built with amorphous cores (*75*). With manufacturing plants in Japan and the United States, Hitachi Metals currently has a global production capacity of 100 million kg per year and dominates this market. The U.S. Department of Energy revealed that in the long term, amorphous cores are likely to predominate in the transformer market because of their higher efficiency (*76*). Asia, for example, has widely adopted amorphous core distribution transformers for the reduction of electrical grid losses. By the end of 2010, the total capacity of these transformers had reached 70 million kVA in China and 35 million kVA in India (*77*). Meanwhile, the worldwide production rate of nanocrystalline materials has exceeded 1 million kg per year and continues to increase (*53*). Researchers are also continually working to improve the performance of both amorphous and nanocrystalline alloys, seeking to boost M_s in particular (*78*).

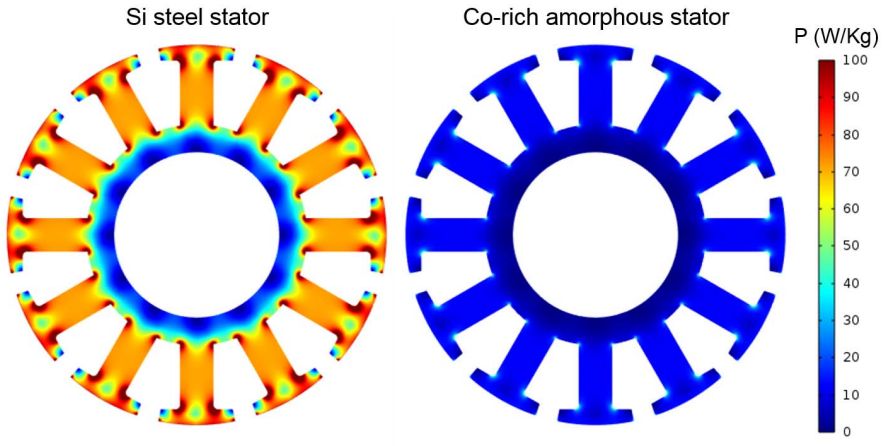


Fig. 9. Motor stator power loss. Distribution of power loss (*P*) in a motor stator fabricated using silicon steel (left) and Co-rich amorphous alloy (right). The motor speed was 18,000 rpm (f = 2.1 kHz). [Reproduced from (*68*)]

Soft magnetic composites

During the past three decades, most of the effort in magnetic composites has focused on the development of powder cores, also known as SMCs. SMCs have been constructed of micrometer-scale particles (most often Fe but sometimes alloys such as MPP, Fe-P, Fe-Si, or Fe-Co), although only very recently have magnetic composite researchers begun taking advantage of nanoparticles. To date, there are no commercially available soft magnetic cores made up of nanoparticles. However, the previous work in SMCs has paved the way for the introduction of nanoparticle-based magnetic composites.

The advantages of composite magnetic cores generally take the form of reduced energy losses during magnetic cycling, which can be strongly affected by composite structure (*36*). By embedding the conductive magnetic particles in an insulating matrix, composites increase ρ (10^{-3} to 10^{-1} microohm·m) and reduce eddy current losses (*79*), although magnetic hysteresis losses systematically increase (*80*). This increase in H_c is caused by the retention of mechanical stresses in both mechanically milled and chemically synthesized particles (*36*). Although a high-temperature anneal could relieve many of these stresses, it would lead to decomposition of the organic binder. This is one area where transitioning to nanometer-sized particles could benefit composites, as it is much easier to anneal away defects and stresses in nanoparticles (*81*).

Magnetic composites also yield an opportunity to tune the μ_r and, in turn, the saturation behavior of the material. In almost all cases, saturation of the magnetic core during operation is highly undesirable. Traditional designs avoid saturation by limiting the applied current, reducing the number of windings, increasing the core size, or introducing an air gap in the magnetic core to reduce μ_r. Similar to gapped cores, spacings of nonmagnetic material in the direction of magnetization can be considered a distributed air gap. A distributed gap avoids the magnetic flux leakage that occurs in a core with a discrete air gap. The gaps in a composite material are exceedingly small because of their large number, and so only a negligible magnetic flux extends beyond the core. A composite (micrometer or nanoscale) magnetic core can be viewed as a distributed air gap inductor with 10 to 10^8 air gaps/cm, depending on the dimensions of the magnetic particulate. In addition to increasing the saturation field, a distributed air gap promotes soft saturation. This condition leads to a slow decrease in μ_r as saturation is approached, making for a less precipitous and disruptive saturation.

Some additional advantages of composite magnetic materials come from added flexibility in manufacturing. Composite materials can be formed with the application of relatively mild forces and low temperatures. For example, an iron powder SMC core is typically made with mild compaction and moderate heating: generally less than 200°C for organic matrices (*82*) and around 500°C for inorganic matrices (*79*). This processing is also amenable to net-shaping, allowing for complex shapes to be produced directly during the compaction stage and without the need for additional machining. Composites also suppress eddy currents isotropically in all directions, leading to additional freedom in the design of systems (*36*).

There is an opportunity to further increase the performance of magnetic composites by transitioning to nanoparticle-based magnetic materials. However, at this point, there have been very few published reports in this area. Transitioning from micro- to nanoscale particles can leverage decades of continuous study of isolated magnetic nanoparticles (*83*). Figure 10 illustrates some of the differences between composites fabricated using micrometer-sized particles and nanoparticles. In some respects, nanocomposites are a continuation of trends noted for larger-scale materials. For example, through the reduction of particle size, eddy current losses can be reduced to the point of being negligible. However, the reduction of particle size also introduces new physics.

Hysteresis losses have a complex relation to particle size. At bulk sizes, soft magnetic materials have modest hysteretic losses, but as their size decreases, the stresses and surface defects present in micrometer-scale materials enhance the hysteresis. Further decreases in scale that approach the characteristic magnetic domain size of a magnetic material lead to a maximum in hysteresis (*84*). Single-domain particles do not reverse their magnetization by domain wall motion but rather through rotation of the magnetic moments within each particle. The energy required for this rotation in the largest single-domain particles is very high. However, this energy decreases linearly with the number of electron spins and, therefore, as r^3 (where r is the particle radius). When the nanoparticles reach a critical size, generally in the tens of nanometers, the thermal energy present in a system provides enough energy to freely reorient the magnetization direction of the particle (*85*). The temperature above which this is true is the superparamagnetic blocking temperature (T_B). These particles no longer exhibit magnetic hysteresis and are referred to as superparamagnetic, as the equations that govern paramagnetism can also be applied, albeit with a much larger number of electron spins (*86*). The elimination of H_c is coupled with the ability to achieve very high values of μ_r.

The weakness of superparamagnetic nanocomposites is the appreciable volume that must be dedicated to the nonmagnetic matrix isolating the individual nanoparticles. If magnetic nanoparticles come into close contact, the particles can magnetically couple and form ferromagnetic domains, leading to domain walls and magnetic hysteresis. Although SMCs with micrometer-sized particles can have very high volumetric loading of magnetic material, it is hard to conceive of a superparamagnetic nanocomposite with much higher than 50% magnetic material.

Although the details of optimization will be complex, by controlling nanoparticle size and spacing, a superparamagnetic nanocomposite can be optimized for use at a specific temperature and frequency. The result would be a material of extremely low loss, with a magnetization that, although lower than that of other magnetic composites, is optimum for a nanocomposite. Because of the complexity of this approach, optimized systems have not been reported to date, though some experiments are beginning to approach this level of control (*87*), and at least one optimization system has been proposed (*88*).

Outlook

Soft magnetic materials have been advancing since the days of Faraday, although more in a series of bursts as opposed to a steady progression. The introduction of WBG semiconductors and new power electronics leveraging them has presented the magnetic materials community with a daunting challenge. WBG devices are also finding their way into faster controllers for electric machines, setting the stage for high rotational

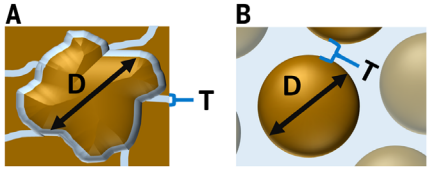


Fig. 10. Magnetic composites. (**A**) Micrometer-scale composites are highly compressed to yield a dense material, typically with irregularly shaped particles. Increasing the particle effective diameter (D) tends to increase M_s and χ while also increasing eddy current losses. As the thickness (T) of the dielectric layer increases, eddy current losses tend to decrease, as do M_s and χ. (**B**) Nanocomposites tend to retain their initial particle shapes and have few defects or residual stress. The interplay between particle diameter (D), spacing (T), M_s, and χ is more complex than in micrometer-scale composites (see discussion in text) but offers the opportunity for more precise tuning of magnetic response. Eddy current losses can become negligible in a well-designed nanocomposite because of the small size of the electrically isolated particles.

speeds and increased efficiency if the development of advanced, high-performance soft magnets can keep up. Many in the research community are now awakening to the need to push research in soft magnetic materials onward with greater urgency.

There are promising magnetic materials that can work at the higher frequencies required in modern power electronics and electrical machines that have been covered in this Review. As mentioned earlier, our coverage was not exhaustive but intended to highlight the most-promising materials for a majority of the high-frequency applications. There is no single soft magnetic material that can satisfy the needs of all power electronic and electrical machine applications. Instead, designers will need to choose judiciously from the available materials, with cost being weighed alongside performance metrics.

Despite being relatively old, soft ferrites continue to be an excellent choice in applications where high power densities are not required. As insulators, they are one of the best materials for mitigating losses, and they are extremely affordable. Even though ferrites have been around since the 1940s, there are still opportunities for improvement through grain-boundary engineering and new processing approaches. Today, amorphous and nanocrystalline alloys are the current state-of-the-art materials. Their distinctive nanostructure coupled with extremely thin laminations keep losses very low while still preserving respectable values of M_s. Additionally, researchers are continuing to steadily improve the M_s of nanocrystalline and amorphous alloys even further. However, the specialized processing required to produce tapes of these materials keeps their cost high. The thin, brittle laminations can also make fabrication of parts challenging.

Technical issues are not the only barrier limiting the adoption of amorphous and nanocrystalline materials to high-end niche applications. Many large and costly electrical machines and transformers still have a long remaining lifetime. Higher up-front costs raise the ratio of capital expenditures to operating expenses, and status quo bias and lack of information affect decision-making. SMCs are proving useful as a relatively low-cost material in rotating electrical machines, due to their isotropic nature. There is hope for much higher performance SMCs, making eddy current and hysteresis losses negligible, by transitioning from micrometer- to nanometer-sized particles. Nanocomposites could also open the door for three-dimensional printing and microfabrication of inductive components, but work on magnetic nanocomposites is only in its initial stages.

The development of entirely new materials and composites is certainly a possibility, and some activity is already occurring in the scientific community (*89*). To lessen barriers to magnetization reversal, it will be necessary to improve control over the nanostructure of both existing and emerging materials. Nanostructure optimization and grain-boundary engineering can also further mitigate eddy current losses. Optimization of the structure within magnetic materials will require increased use of advanced characterization tools. These methodologies will include techniques applied to almost all materials, such as electron microscopy and x-ray (including synchrotron-based) characterization. Techniques specific to imaging magnetic domain structure—such as Lorentz microscopy (*90*), electron holography (*91*), magnetic force microscopy (*92*), x-ray magnetic circular dichroism spectroscopy (*93*), and imaging based on MOKE (*94*, *95*)—will be particularly helpful. It is now up to the community of materials scientists, physicists, chemists, and others to apply these techniques and our knowledge of magnetic materials to meet the demands of the next generation of power electronics and electrical machines.

REFERENCES AND NOTES

1. F. Fiorillo, G. Bertotti, C. Appino, M. Pasquale, "Soft magnetic materials" in *Wiley Encyclopaedia of Electrical and Electronics Engineering*, J. G. Webster, Ed. (Wiley, 2016).
2. F. Fiorillo, *Measurement and Characterization of Magnetic Materials* (Elsevier, 2004).
3. H. J. Williams, W. Shockley, C. Kittel, Studies of the Propagation Velocity of a Ferromagnetic Domain Boundary. *Phys. Rev.* **80**, 1090–1094 (1950). doi: 10.1103/PhysRev.80.1090
4. R. H. Pry, C. P. Bean, Calculation of the Energy Loss in Magnetic Sheet Materials Using a Domain Model. *J. Appl. Phys.* **29**, 532–533 (1958). doi: 10.1063/1.1723212
5. G. Bertotti, *Hysteresis in Magnetism* (Academic Press, 1998)
6. S. Flohrer *et al.*, Dynamic magnetization process of nanocrystalline tape wound cores with transverse field-induced anisotropy. *Acta Mater.* **54**, 4693–4698 (2006). doi: 10.1016/j.actamat.2006.04.040
7. G. Bertotti, Physical interpretation of eddy current losses in ferromagnetic materials. I. Theoretical considerations. *J. Appl. Phys.* **57**, 2110–2117 (1985). doi: 10.1063/1.334404
8. G. Bertotti, General properties of power losses in soft ferromagnetic materials. *IEEE Trans. Magn.* **24**, 621–630 (1988). doi: 10.1109/20.43994
9. F. Fiorillo, Measurements of magnetic materials. *Metrologia* **47**, S114–S142 (2010). doi: 10.1088/0026-1394/47/2/S11
10. F. Fiorillo, E. Ferrara, M. Coïsson, C. Beatrice, N. Banu, Magnetic properties of soft ferrites and amorphous ribbons up to radiofrequencies. *J. Magn. Magn. Mater.* **322**, 1497–1504 (2010). doi: 10.1016/j.jmmm.2009.09.038
11. F. Fiorillo, C. Appino, M. Pasquale, in *The Science of Hysteresis Vol. II*, G. Bertotti, I. Mayergoyz, Eds. (Academic Press, 2006), pp. 1–189.
12. A. Magni *et al.*, Domain wall processes, rotations, and high-frequency losses in thin laminations. *IEEE Trans. Magn.* **48**, 3796–3799 (2012). doi: 10.1109/TMAG.2012.2196985
13. J. Popović-Gerber *et al.*, Power Electronics Enabling Efficient Energy Usage: Energy Savings Potential and Technological Challenges. *IEEE Trans. Power Electron.* **27**, 2338–2353 (2012). doi: 10.1109/TPEL.2011.2171195
14. B. Kroposki *et al.*, Benefits of Power Electronic Interfaces for Distributed Energy Systems. *IEEE Trans. Energ. Convers.* **25**, 901–908 (2010). doi: 10.1109/TEC.2010.2053975
15. B. K. Bose, Global Warming: Energy, Environmental Pollution, and the Impact of Power Electronics. *IEEE Ind. Electron. Mag.* **4**, 6–17 (2010). doi: 10.1109/MIE.2010.935860
16. D. Tiku, dc Power Transmission: Mercury-Arc to Thyristor HVdc Valves. *IEEE Power Energy Mag.* **12**, 76–96 (2014) [History]. doi: 10.1109/MPE.2013.2293398
17. F. Iacopi, M. Van Hove, M. Charles, K. Endo, Power electronics with wide bandgap materials: Toward greener, more efficient technologies. *MRS Bull.* **40**, 390–395 (2015). doi: 10.1557/mrs.2015.71
18. S. Dimitrijev, J. Han, H. A. Moghadam, A. Aminbeidokhti, Power-switching applications beyond silicon: Status and future prospects of SiC and GaN devices. *MRS Bull.* **40**, 399–405 (2015). doi: 10.1557/mrs.2015.89
19. S. Bhattacharya, Transforming the transformer. *IEEE Spectr.* **54**, 38–43 (2017). doi: 10.1109/MSPEC.2017.7951721
20. S. Krishnamurthy, in *2012 Twenty-Seventh Annual IEEE Applied Power Electronics Conference and Exposition (APEC)* (IEEE, 2012), pp. 1414–1417.
21. J. W. Kolar, J. Biela, S. Waffler, T. Friedli, U. Badstuebner, "Performance trends and limitations of power electronic systems" in *2010 6th International Conference on Integrated Power Electronics Systems* (IEEE, 2010), pp. 1–20.
22. P. Waide, C. U. Brunner, "Energy-efficiency policy opportunities for electric motor-driven systems," International Energy Agency (IEA) working paper (2011); www.iea.org/publications/freepublications/publication/EE_for_ElectricSystems.pdf.
23. A. Krings, A. Boglietti, A. Cavagnino, S. Sprague, Soft magnetic material status and trends in electric machines. *IEEE Trans. Ind. Electron.* **64**, 2405–2414 (2017). doi: 10.1109/TIE.2016.2613844
24. G. A. McCoy, T. Litman, J. G. Douglass, "Energy-efficient electric motor selection handbook" (Bonneville Power Administration, 1990); https://doi.org/10.2172/6116458.
25. A. Krings, thesis, KTH, Stockholm (2014).
26. J. M. Silveyra, A. Leary, V. DeGeorge, S. Simizu, M. McHenry, High speed electric motors based on high performance novel soft magnets. *J. Appl. Phys.* **115**, 17A319 (2014). doi: 10.1063/1.4864247
27. B. D. Cullity, C. D. Graham, *Introduction to Magnetic Materials* (IEEE Press, ed. 2nd, 2009)
28. "Super Core™ electrical steel sheets for high-frequency application" (JFE Steel Corporation, 2017); www.jfe-steel.co.jp/en/products/electrical/catalog/f1e-002.pdf.
29. K. Shoji, N. Misao, H. Tatsuhiko, "Recent Progress of High Silicon Electrical Steel in JFE Steel" (JFE Technical Report, JFE Steel Corporation, 2016); www.jfe-steel.co.jp/en/research/report/021/pdf/021-04.pdf.
30. H. D. Arnold, G. W. Elmen, Permalloy, A New Magnetic Material of Very High Permeability. *Bell Syst. Tech. J.* **2**, 101–111 (1923). doi: 10.1002/j.1538-7305.1923.tb03595.x
31. A. Alabakhshizadeh, O. M. Midtgård, K. Boysen, in *2013 IEEE 39th Photovoltaic Specialists Conference (PVSC)* (IEEE, 2013), pp. 2845–2848.
32. "MPP Cores," Magnetics®, www.mag-inc.com/Products/Powder-Cores/MPP-Cores.
33. M. J. de Vries, *80 Years of Research at the Philips Natuurkundig Laboratorium (1914-1994). The Role of the Nat. Lab. at Philips* (Amsterdam Univ. Press, 2005).
34. P. Duwez, S. C. H. Lin, Amorphous ferromagnetic phase in iron-carbon-phosphorus alloys. *J. Appl. Phys.* **38**, 4096–4097 (1967). doi: 10.1063/1.1709084
35. Y. Yoshizawa, S. Oguma, K. Yamauchi, New Fe-based soft magnetic alloys composed of ultrafine grain structure. *J. Appl. Phys.* **64**, 6044–6046 (1988). doi: 10.1063/1.342149
36. H. Shokrollahi, K. Janghorban, Soft magnetic composite materials (SMCs). *J. Mater. Process. Technol.* **189**, 1–12 (2007). doi: 10.1016/j.jmatprotec.2007.02.034

37. A. G. Jack *et al*., Permanent-magnet machines with powdered iron cores and prepressed windings. *IEEE Trans. Ind. Appl.* **36**, 1077–1084 (2000). doi: 10.1109/28.855963
38. L. O. Hultman, A. G. Jack, in *Electric Machines and Drives Conference, 2003. IEMDC'03* (IEEE, 2003), pp. 516–522.
39. L. Néel, Propriétés magnétiques des ferrites: Ferrimagnétisme et antiferromagnétisme. *Ann. Phys.* **12**, 137–198 (1948). doi: 10.1051/anphys/194812030137
40. P. W. Anderson, Antiferromagnetism. Theory of Superexchange Interaction. *Phys. Rev.* **79**, 350–356 (1950). doi: 10.1103/PhysRev.79.350
41. All materials have a magnetic moment and can be described as diamagnetic, paramagnetic, ferromagnetic, etc. Because diamagnetic and paramagnetic materials are so weakly magnetic that their contribution to a ferromagnetic or ferrimagnetic material can be ignored, they will be referred to as nonmagnetic for simplicity.
42. L. K. V. F. Mazaleyrat, L. K. Varga, Ferromagnetic composites. *J. Magn. Magn. Mater.* **215–216**, 253–259 (2000). doi: 10.1016/S0304-8853(00)00128-1
43. J. L. Snoek, Dispersion and absorption in magnetic ferrites at frequencies above one Mc/s. *Physica* **14**, 207–217 (1948). doi: 10.1016/0031-8914(48)90038-X
44. P. R. Ohodnicki *et al*., Correlation between texture, anisotropy, and vector magnetization processes investigated by two-dimensional vector vibrating sample magnetometry in BaO(Fe2O3)6 thin film. *J. Appl. Phys.* **103**, 07E514 (2008). doi: 10.1063/1.2838630
45. Y. Chen, V. G. Harris, U.S. patent 9,117,565, Magnetic grain boundary engineered ferrite core materials (2015).
46. A. W. Roesler *et al*., Planar LTCC Transformers for High-Voltage Flyback Converters. *IEEE Trans. Compon. Packag. Tech.* **33**, 359–372 (2010). doi: 10.1109/TCAPT.2009.2031872
47. R. Alben, J. Becker, M. Chi, Random anisotropy in amorphous ferromagnets. *J. Appl. Phys.* **49**, 1653–1658 (1978). doi: 10.1063/1.324881
48. G. Herzer, Grain structure and magnetism of nanocrystalline ferromagnets. *IEEE Trans. Magn.* **25**, 3327–3329 (1989). doi: 10.1109/20.42292
49. G. Herzer, "Soft magnetic materials—Nanocrystalline alloys" in *Handbook of Magnetism and Advanced Magnetic Materials* (Wiley, 2007); http://doi.org/10.1002/9780470022184.hmm402.
50. A. M. Leary, P. R. Ohodnicki, M. E. McHenry, Soft magnetic materials in high-frequency, high-power conversion applications. *J. Minerals Metals Mater. Soc.* **64**, 772–781 (2012). doi: 10.1007/s11837-012-0350-0
51. V. DeGeorge, S. Shen, P. Ohodnicki, M. Andio, M. E. McHenry, Multiphase Resistivity Model for Magnetic Nanocomposites Developed for High Frequency, High Power Transformation. *J. Electron. Mater.* **43**, 96–108 (2014). doi: 10.1007/s11664-013-2835-1
52. L. K. Varga, É. Bakos, É. Kisdi-Koszó, É. Zsoldos, L. F. Kiss, Time and temperature dependence of nanocrystalline structure formation in a Finemet-type amorphous alloy. *J. Magn. Magn. Mater.* **133**, 280–282 (1994). doi: 10.1016/0304-8853(94)90546-0
53. G. Herzer, Modern soft magnets: Amorphous and nanocrystalline materials. *Acta Mater.* **61**, 718–734 (2013). doi: 10.1016/j.actamat.2012.10.040
54. M. Willard *et al*., Structure and magnetic properties of (Fe 0.5 Co 0.5) 88 Zr 7 B 4 Cu 1 nanocrystalline alloys. *J. Appl. Phys.* **84**, 6773–6777 (1998). doi: 10.1063/1.369007
55. A. Leary *et al*., Stress induced anisotropy in Co-rich magnetic nanocomposites for inductive applications. *J. Mater. Res.* **31**, 3089–3107 (2016). doi: 10.1557/jmr.2016.324
56. J. Turčanová *et al*., Magnetic and mechanical properties of nanocrystalline Fe-Ni-Nb-B Alloys. *J. Phys. Conf. Ser.* **144**, 012065 (2009). doi: 10.1088/1742-6596/144/1/012065
57. N. Aronhime, V. DeGeorge, V. Keylin, P. Ohodnicki, M. E. McHenry, The Effects of Strain-Annealing on Tuning Permeability and Lowering Losses in Fe-Ni-Based Metal Amorphous Nanocomposites. *JOM* **69**, 2164–2170 (2017). doi: 10.1007/s11837-017-2480-x
58. J. May, C. A. C. Souza, C. Morelli, N. Mariano, S. E. Kuri, Magnetic and corrosion properties comparison of FeSi-based, FeZr-based and FeCo-based alloys. *J. Alloys Compd.* **390**, 106–111 (2005). doi: 10.1016/j.jallcom.2004.05.090
59. C. Souza, D. Ribeiro, C. Kiminami, Corrosion resistance of Fe-Cr-based amorphous alloys: An overview. *J. Non-Cryst. Solids* **442**, 56–66 (2016). doi: 10.1016/j.jnoncrysol.2016.04.009
60. J. M. Silveyra, E. Illeková, Effects of air annealing on Fe–Si–B–M–Cu (M= Nb, Mo) alloys. *J. Alloys Compd.* **610**, 180–183 (2014). doi: 10.1016/j.jallcom.2014.04.147
61. J. M. Silveyra *et al*., High performance of low cost soft magnetic materials. *Bull. Mater. Sci.* **34**, 1407–1413 (2011). doi: 10.1007/s12034-011-0336-5
62. A. Makino, H. Men, T. Kubota, K. Yubuta, A. Inoue, FeSiBPCu nanocrystalline soft magnetic alloys with high Bs of 1.9 Tesla produced by crystallizing hetero-amorphous phase. *Mater. Trans.* **50**, 204–209 (2009). doi: 10.2320/matertrans.MER2008306
63. A. D. Setyawan *et al*., Magnetic properties of 120-mm wide ribbons of high Bs and low core-loss NANOMET® alloy. *J. Appl. Phys.* **117**, 17B715 (2015). doi: 10.1063/1.4913936
64. W. Mischler, G. Rosenberry, P. Frischmann, R. Tompkins, Test results on a low loss amorphous iron induction motor. *IEEE Trans. Power Apparatus Syst.* **PAS-100**, 2907–2911 (1981). doi: 10.1109/TPAS.1981.316408
65. Z. Wang *et al*., Development of a permanent magnet motor utilizing amorphous wound cores. *IEEE Trans. Magn.* **46**, 570–573 (2010). doi: 10.1109/TMAG.2009.2033350
66. T. Fukao, A. Chiba, M. Matsui, Test results on a super-high-speed amorphous-iron reluctance motor. *IEEE Trans. Ind. Appl.* **25**, 119–125 (1989). doi: 10.1109/28.18881
67. R. Kolano, A. Kolano-Burian, M. Polak, J. Szynowski, Application of rapidly quenched soft magnetic materials in energy-saving electric equipment. *IEEE Trans. Magn.* **50**, 1–4 (2014). doi: 10.1109/TMAG.2013.2283918
68. J. M. Silveyra *et al*., Amorphous and nanocomposite materials for energy-efficient electric motors. *J. Electron. Mater.* **45**, 219–225 (2016). doi: 10.1007/s11664-015-3968-1
69. S. Okamoto, N. Denis, Y. Kato, M. Ieki, K. Fujisaki, Core loss reduction of an interior permanent-magnet synchronous motor using amorphous stator core. *IEEE Trans. Ind. Appl.* **52**, 2261–2268 (2016). doi: 10.1109/TIA.2016.2532279
70. W. Tong, S. Wu, J. Sun, L. Zhu, in *Vehicle Power and Propulsion Conference (VPPC)* (IEEE, 2016); https://doi.org/10.1109/VPPC.2016.7791716.
71. T. Yamazakii, H. Uchiyama, K. Nakazawa, T. Isomura, H. Ogata, "The development of direct drive motors for solar cars" SAE Technical Paper 2017-01-1232 (2017); https://doi.org/10.4271/2017-01-1232.
72. Z. Wang, Y. Enomoto, H. Tokoi, A. Komura, T. Obata, K. Souma, "Development of IE5 high efficiency motor with iron-base amorphous magnetic cores," paper presented at the Energy Efficiency in Motor Driven Systems (EEMODS) '15 Conference, Helsinski, Finland, 16 September 2015.
73. N. Nishiyama, K. Tanimoto, A. Makino, Outstanding efficiency in energy conversion for electric motors constructed by nanocrystalline soft magnetic alloy "NANOMET®" cores. *AIP Adv.* **6**, 055925 (2016). doi: 10.1063/1.4944341
74. N. Denis, M. Inoue, K. Fujisaki, H. Itabashi, T. Yano, Iron loss reduction of permanent magnet synchronous motor by use of stator core made of nanocrystalline magnetic material. *IEEE Trans. Magn.* **53**, 8110006 (2017). doi: 10.1109/TMAG.2017.2700471
75. "Amorphous Metal Distribution Transformers" (Hitachi Metals America, (2016); http://metglas.com/wp-content/uploads/2016/12/Metglas-Power-Brochure.pdf.
76. "Energy Conservation Program: Energy Conservation Standards for Distribution Transformers. Final rule," (Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy, 2013); https://www1.eere.energy.gov/buildings/appliance_standards/pdfs/dt_final_rule.pdf.
77. "Review and the future of amorphous metal transformers in Asia" (Project4 Media, 2011); https://web.archive.org/web/20120406141302/http://www.amorphous-metal-transformer.com/ufiles/fck/file/Book%201_new_20110907.pdf.
78. E. Theisen, Development of New Amorphous and Nanocrystalline Magnetic Materials for Use in Energy-Efficient Devices. *MRS Adv.* **2**, 3409–3414 (2017). doi: 10.1557/adv.2017.552
79. A. H. Taghvaei, H. Shokrollahi, K. Janghorban, Properties of iron-based soft magnetic composite with iron phosphate–silane insulation coating. *J. Alloys Compd.* **481**, 681–686 (2009). doi: 10.1016/j.jallcom.2009.03.074
80. M. Anhalt, Systematic investigation of particle size dependence of magnetic properties in soft magnetic composites. *J. Magn. Magn. Mater.* **320**, e366–e369 (2008). doi: 10.1016/j.jmmm.2008.02.072
81. D. Guo, G. Xie, J. Luo, Mechanical properties of nanoparticles: Basics and applications. *J. Phys. D* **47**, 013001 (2014). doi: 10.1088/0022-3727/47/1/013001
82. A. H. Taghvaei, H. Shokrollahi, K. Janghorban, H. Abiri, Eddy current and total power loss separation in the iron–phosphate–polyepoxy soft magnetic composites. *Mater. Des.* **30**, 3989–3995 (2009). doi: 10.1016/j.matdes.2009.05.026
83. J. L. Dormann, Le phénomène de superparamagnétisme. *Rev. Phys. Appl.* **16**, 275–301 (1981). doi: 10.1051/rphysap:01981001606027500
84. I. Jacobs, C. Bean, in *Magnetism*, George T. Rado, Harry Suhl, Eds. (Academic Press, 1963), vol. 3, pp. 271–350.
85. L. Neel, Influence des Fluctuations Thermiques sur Laimantation de Grains Ferromagnetiques Tres Fins. *C. R. Hebd. Seances Acad. Sci.* **228**, 664–666 (1949).
86. C. Bean, J. Livingston, Superparamagnetism. *J. Appl. Phys.* **30**, 120S–129S (1959). doi: 10.1063/1.2185850
87. H. Yun *et al*., Alternate current magnetic property characterization of nonstoichiometric zinc ferrite nanocrystals for inductor fabrication via a solution based process. *J. Appl. Phys.* **119**, 113901 (2016). doi: 10.1063/1.4942865
88. R. J. Kaplar, J. C. Neely, D. L. Huber, L. J. Rashkin, Generation-After-Next Power Electronics: Ultrawide-bandgap devices, high-temperature packaging, and magnetic nanocomposite materials. *IEEE Power Electron. Mag.* **4**, 36–42 (2017). doi: 10.1109/MPEL.2016.2643098
89. T. C. Monson, E. J. Lavernia, B. Zheng, Y. Zhou, "Method to Synthesize Bulk Iron Nitride," U.S. Patent 9,963,344 B2 (2018).
90. T. Matsuda, K. Harada, H. Kasai, O. Kamimura, A. Tonomura, Observation of dynamic interaction of vortices with pinning centers by Lorentz microscopy. *Science* **271**, 1393–1395 (1996). doi: 10.1126/science.271.5254.1393
91. H. Lichte, M. Lehmann, Electron holography—basics and applications. *Rep. Prog. Phys.* **71**, 016102 (2008). doi: 10.1088/0034-4885/71/1/016102
92. D. Rugar *et al*., Magnetic force microscopy: General principles and application to longitudinal recording media. *J. Appl. Phys.* **68**, 1169–1183 (1990). doi: 10.1063/1.346713
93. P. Fischer *et al*., Imaging of magnetic domains by transmission x-ray microscopy. *J. Phys. D* **31**, 649–655 (1998). doi: 10.1088/0022-3727/31/6/012
94. R. Schaefer, G. Herzer, Continuous magnetization patterns in amorphous ribbons. *IEEE Trans. Magn.* **37**, 2245–2247 (2001). doi: 10.1109/20.951137
95. J. McCord, Progress in magnetic domain observation by advanced magneto-optical microscopy. *J. Phys. D* **48**, 333001 (2015). doi: 10.1088/0022-3727/48/33/333001
96. L. Dobrzański, M. Drak, B. Ziębowicz, Materials with specific magnetic properties. *J. Achiev. Mater. Manufact. Eng.* **17**, 37–40 (2006).

ACKNOWLEDGMENTS

E.F. thanks F. Fiorillo for support and suggestions and T.C.M. thanks I. Gyuk, director of the U.S. Department of Energy (DOE) Energy Storage program. **Funding:** J.M.S. acknowledges support from CONICET and UBACyT 20020150200168BA. T.C.M. and D.L.H. were funded in part by Sandia's Laboratory Directed Research and Development program. T.C.M. acknowledges partial support from the Energy Storage program of the DOE, Office of Electricity Delivery and Energy Reliability. D.L.H. acknowledges partial support from the Center for Integrated Nanotechnologies, an Office of Science User Facility operated for the DOE Office of Science. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the DOE's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States government. **Competing interests:** None declared.

10.1126/science.aao0195

RESEARCH ARTICLE SUMMARY

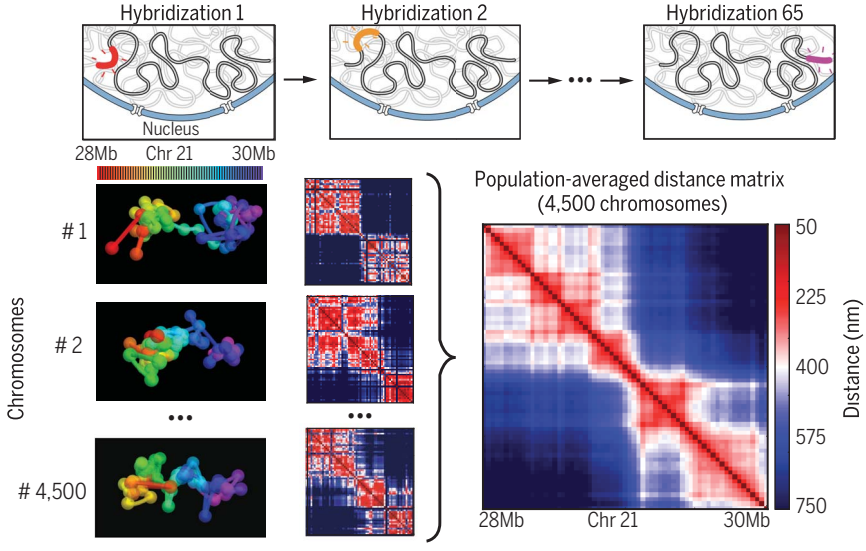
CHROMATIN STRUCTURE

Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells

Bogdan Bintu*, Leslie J. Mateo*, Jun-Han Su, Nicholas A. Sinnott-Armstrong, Mirae Parker, Seon Kinrot, Kei Yamaya, Alistair N. Boettiger†‡, Xiaowei Zhuang†‡

INTRODUCTION: Chromatin adopts an intricate three-dimensional (3D) organization in the nucleus that is critical for many genome processes, from gene regulation to genome replication. Our understanding of the chromatin organization remains relatively poor at the kilobase-to-megabase scale, which spans the sizes of individual genes and regulatory domains and is thus of critical importance to genome regulation. Recent Hi-C experiments revealed topologically associating domains (TADs) as a ubiquitous chromatin organization feature in many organisms. However, the basic properties of TADs, including whether TADs represent a fundamental unit of genome organization in individual cells or an emergent property from cell population averaging, and the formation mechanism of TADs, remain unclear. In addition, our understanding of genome organization is largely built on pairwise interactions, whereas relatively little is known about higher-order chromatin interactions. Methods that provide a high-resolution visualization of chromatin structure in individual cells will elucidate these and many other questions related to genome organization.

RATIONALE: We report a super-resolution chromatin tracing method that allows determination of both the structural features and their genomic coordinates with high resolution in single cells. We reason that if numerous chromatin loci could be identified and precisely localized in individual cells, connecting their positions would allow us to trace the chromatin conformation. However, typical multicolor imaging allows only a few loci to be simultaneously imaged. To overcome this challenge, we partitioned the targeted genomic region into numerous segments, each 30 kb in length, and imaged individual segments using sequential rounds of fluorescence in situ hybridization. This allowed us to generate a 3D super-resolution image of the chromatin in numerous pseudocolors, each reporting the position and structure of a 30-kb segment with nanometer-scale precision.



Super-resolution chromatin tracing reveals TAD-like domain structures in single cells. Consecutive 30-kb segments of a chromatin region of interest were sequentially imaged with diffraction-limited or super-resolution fluorescence microscopy. The pseudocolored images of the positions of individual chromatin segments in single cells and the corresponding matrices of intersegment distances reveal TAD-like structures with a globular conformation in single cells. The population-average matrix reveals TADs at the ensemble level.

RESULTS: Our imaging data revealed an abundance of TAD-like domain structures with spatially segregated globular conformations in single cells. The domain boundaries varied from cell to cell, exhibiting nonzero probability of residing at any genome positions, but with a preference at CCCTC-binding factor (CTCF)- and cohesin-binding sites. Notably, cohesin depletion, which abolished TADs at the population-average level, did not alter the prevalence of TAD-like structures in single cells; only the preferential positioning of domain boundaries was lost, explaining the loss of population-level TADs. Our results suggest that cohesin is not required for the formation or maintenance of single-cell domain structures, but that their preferential boundary positions are influenced by cohesin-CTCF interaction.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aau1783

In addition, we observed prevalent multiway interactions among triplets of chromatin loci. These higher-order interactions were cooperative, i.e., most three-way contacts were observed at higher frequencies than would be expected from the frequency of pairwise interactions. Notably, these multiway interactions were also retained after cohesin depletion.

CONCLUSION: Our imaging method offers a high-resolution physical view of chromatin conformation of targeted genomic regions in single cells, providing a powerful and complementary approach to sequencing-based genome-wide methods for interrogating genome organization. The TAD-like structures and multiway chromatin interactions observed in single cells add important constraints on genome folding and have implications for understanding the role of genome structure in diverse biological processes from enhancer-promoter communication to genome compartmentalization. We envision that future work will further improve the resolution and genomic coverage of this approach and will combine the imaging of chromatin with regulatory factors and/or expressed RNAs to reveal the underlying mechanism and functional implication of chromosome organization. ■

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†These authors contributed equally to this work.
‡Corresponding author. Email: zhuang@chemistry.harvard.edu (X.Z.); boettiger@stanford.edu (A.N.B.)
Cite this article as B. Bintu *et al.*, *Science* **362**, eaau1783 (2018). DOI: 10.1126/science.aau1783

RESEARCH ARTICLE

CHROMATIN STRUCTURE

Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells

Bogdan Bintu[1]*, Leslie J. Mateo[2]*, Jun-Han Su[1], Nicholas A. Sinnott-Armstrong[3], Mirae Parker[4]†, Seon Kinrot[1], Kei Yamaya[2], Alistair N. Boettiger[2]‡§, Xiaowei Zhuang[1]‡§

The spatial organization of chromatin is pivotal for regulating genome functions. We report an imaging method for tracing chromatin organization with kilobase- and nanometer-scale resolution, unveiling chromatin conformation across topologically associating domains (TADs) in thousands of individual cells. Our imaging data revealed TAD-like structures with globular conformation and sharp domain boundaries in single cells. The boundaries varied from cell to cell, occurring with nonzero probabilities at all genomic positions but preferentially at CCCTC-binding factor (CTCF)- and cohesin-binding sites. Notably, cohesin depletion, which abolished TADs at the population-average level, did not diminish TAD-like structures in single cells but eliminated preferential domain boundary positions. Moreover, we observed widespread, cooperative, multiway chromatin interactions, which remained after cohesin depletion. These results provide critical insight into the mechanisms underlying chromatin domain and hub formation.

Three-dimensional (3D) organization of the genome and cis interactions between genomic loci regulate many cellular processes, including gene expression, DNA replication, and DNA damage repair (*1*–*6*). Recent development of chromosome conformation capture technologies, such as Hi-C (*7*), have greatly enriched our understanding of chromatin organization (*4*–*6*), revealing genome-wide structural features, such as topologically associating domains (TADs) and CTCF-dependent chromatin loops (*8*–*12*).

TADs are revealed in ensemble-averaged Hi-C contact maps as domains within which chromatin shows high contact probability (*8*–*11*). TADs tend to coincide with epigenetic domains, harbor co-regulated genes, and are generally conserved across cell types and species (*4*–*6*, *8*–*11*). At a finer scale, TADs are divided into smaller domains with enhanced contact frequency, named sub-TADs (or contact domains), which are more variable across different cell types and thought to be involved in differential gene expression (*6*, *12*, *13*). Despite the proposed central role of these domain structures in chromatin organization and genome function, many of their basic properties remain unclear. In one view, it has been proposed that TADs represent a fundamental physical unit of the genome organization within individual cells, which promote intradomain chromatin interactions but inhibit interdomain interactions through spatial segregation (*6*, *14*, *15*). Recent super-resolution imaging studies provide partial support for this view by showing repressed chromatin domains as spatially segregated compact structures or nanocompartments in single cells (*16*, *17*). However, recently reported single-cell Hi-C maps exhibit only a low density of chromatin contacts in individual cells and cell-to-cell variability in these contacts, leading to the debate over whether TADs exist in single cells (*18*–*21*). Although enrichment of chromatin contacts and, occasionally, large TAD-like structures could be observed in some genomic regions in single-cell Hi-C maps, the sparsity of contacts in these maps makes de novo identification of individual chromatin domains and domain boundaries challenging in single cells. An alternative view has thus been proposed that the genome is not packaged into spatially segregated domain structures in single cells but is largely organized by recurrent pairwise interactions in an otherwise diverse ensemble of conformational configurations, with TADs being considered an emergent property from cell population averaging due to a tendency for contact enrichment within specific genomic regions in individual cells (*19*, *22*). These two distinct models of TADs have different implications for our understanding of cis regulation of chromatin, but unfortunately a clear physical understanding of the TAD structure is still missing. Hi-C experiments have also identified "loop interactions," primarily observed at the boundaries of TADs or contact domains that harbor convergent CCCTC-binding factor (CTCF) sites (*4*–*6*, *12*). Loop interactions have been proposed to facilitate interactions between regulatory sequences near the CTCF sites, such as enhancers and promoters, to induce gene activation (*4*, *12*). Although numerous pairwise loop interactions have been observed in various genomes by Hi-C and other methods, higher-order interactions that involve more than two genomic loci are only beginning to be explored (*23*–*27*).

A major challenge in addressing the questions regarding chromatin organization is the lack of tools to provide a high-resolution visualization of the physical structure of chromatin in individual cells at the kilobase-to-megabase scale, which spans the sizes of genes and regulatory domains. Despite recent innovations in imaging methods that advance our knowledge of chromatin organization at this scale [see for example, (*16*, *17*, *28*–*31*)], current microscopy approaches provide limited sequence information and resolution, and hence the power of high–spatial-resolution visualization is not accompanied by an ability to map the genomic sequences of chromatin structures de novo.

Multiplexed super-resolution fluorescence in situ hybridization (FISH) imaging for chromatin tracing

In this work, we developed a highly multiplexed super-resolution imaging approach for chromatin conformation tracing, which allows unbiased determination of both the structural features and their genomic coordinates with high resolution in single cells. To trace chromatin organization within and across TADs and sub-TADs, we imaged multiple 1.2- to 2.5-Mb regions of human chromosome 21 (Chr 21), traversing different numbers of TADs and sub-TADs, in multiple cell types. We partitioned each region of interest into consecutive 30-kb segments and labeled and imaged individual segments following a sequential hybridization protocol, modified from our previous multiplexed RNA imaging method (*32*) and our previous lower- (megabase) resolution chromatin imaging work (*33*). In the first step, we labeled the entire region with a library of ~12,000 to 25,000 primary Oligopaint probes (*34*, *35*), each primary probe containing a 20-nucleotide (nt) readout sequence that was specific for each 30-kb segment to facilitate multiplexed FISH imaging (Fig. 1A and table S1) (*33*). Next, we added dye-labeled readout probes complementary to the readout sequences to allow three-dimensional (3D) stochastic optical reconstruction microscopy (STORM) (*36*, *37*) or 3D diffraction-limited imaging of individual 30-kb segments (Fig. 1A and table S2). After each round, imaging one or two segments with single- or two-color imaging, the signal of the readout probes was extinguished by using a strand

[1]Howard Hughes Medical Institute, Department of Chemistry and Chemical Biology and Department of Physics, Harvard University, Cambridge, MA 02138, USA. [2]Department of Developmental Biology, Stanford University, Stanford, CA 94305, USA. [3]Department of Genetics, Stanford University, Stanford, CA 94305, USA. [4]Department of Physics, Stanford University, Stanford, CA 94305, USA.
*These authors contributed equally to this work.
†Present address: Department of Computational Systems Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
‡These authors contributed equally to this work.
§Corresponding author. Email: zhuang@chemistry.harvard.edu (X.Z.); boettiger@stanford.edu (A.N.B.)

displacement reaction to remove the readout probes or by using photobleaching, or both, and the sequential process of readout-probe labeling and imaging was repeated until all segments were imaged (Fig. 1A). This allowed us to generate, for each cell, a 3D super-resolution image of the chromatin region of interest in numerous pseudocolors, each reporting the position and structure of a contiguous 30-kb segment (Fig. 1B), with <50-nm error in their localization and <5% error in their physical sizes (fig. S1).

These super-resolution chromatin images allowed us to measure the pairwise interactions between chromatin segments and compare the results with ensemble Hi-C measurements. We first focused on a 1.2-Mb region of Chr 21 (Chr21:28Mb-29.2Mb) in IMR90 fibroblast cells using STORM imaging. Quantitatively, we determined two complementary metrics between each pair of segments from the STORM images: the spatial overlap and the centroid-to-centroid distance (Fig. 1C). For each metric, we constructed a matrix for the entire imaged region for every copy of the chromosome imaged and averaged across ~250 imaged chromosomes to obtain a population view, which can be compared to Hi-C (Fig. 1, D to F). These matrices derived from imaging (Fig. 1, E and F, and fig. S2A) displayed domain structures (block-like

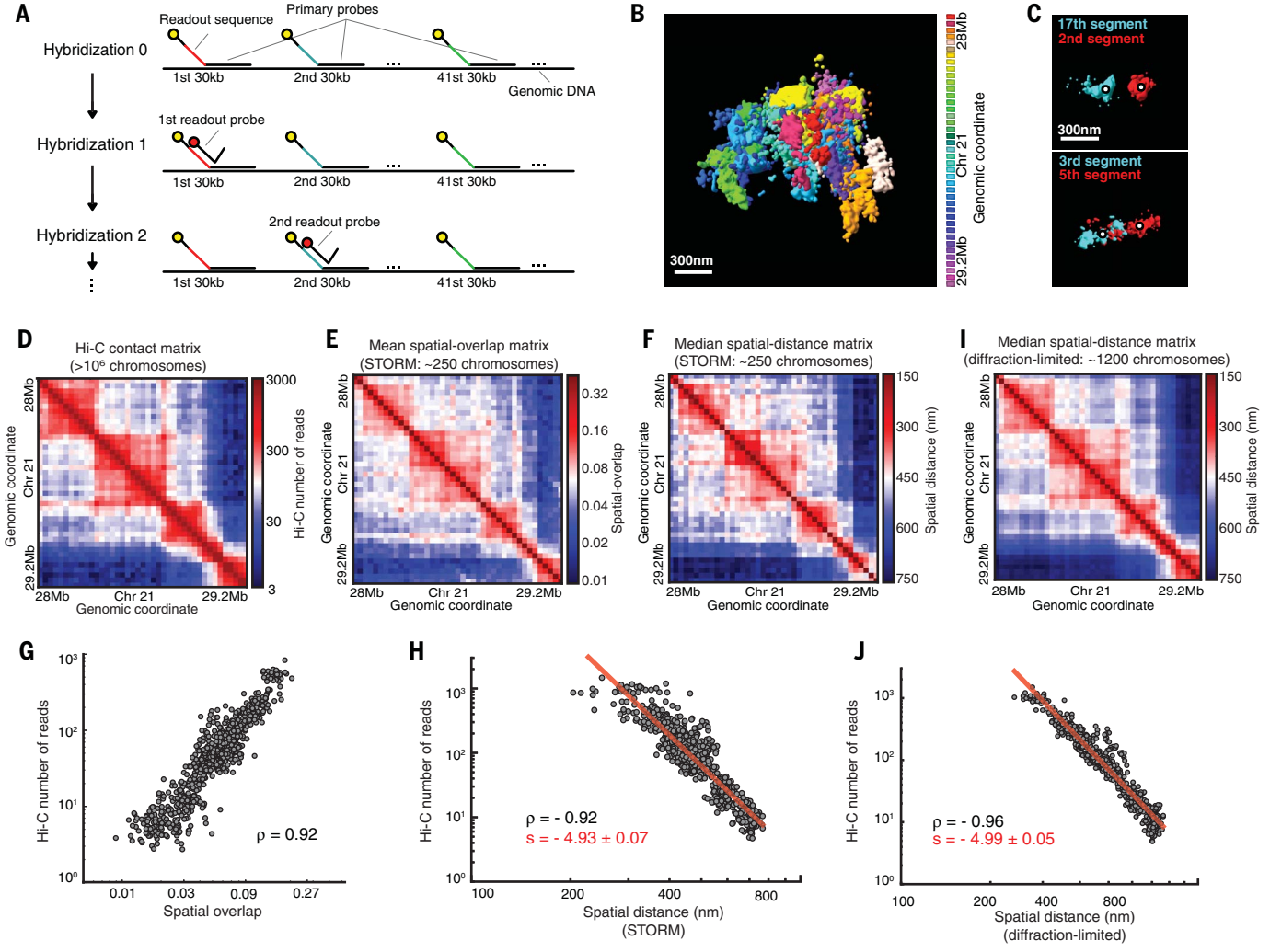


Fig. 1. Multiplex FISH imaging for high-resolution chromatin tracing allows de novo identification of TADs and sub-TADs. (**A**) A scheme of the imaging approach. The genomic region of interest is partitioned into consecutive 30-kb segments and first hybridized with primary oligonucleotide probes that label all segments. These probes contained a readout sequence specific to each 30-kb segment. Each segment is labeled by ~300 probes, but only one is shown. Readout probes complementary to the readout sequences are then added sequentially, allowing the imaging of individual 30-kb segments. (**B**) Composite 3D STORM images of 41 consecutive 30-kb chromatin segments in a 1.2-Mb region of chromosome 21 (Chr21:28Mb-29.2Mb), in 41 pseudocolors, in one copy of Chr 21 of an IMR90 cell. (**C**) 3D STORM images of two pairs of chromatin segments showing different degree of overlap but similar distances between their center positions (marked by white dots). (**D**) Ensemble Hi-C contact frequency matrix for the 1.2-Mb genomic region binned at 30-kb resolution [data from (*12*)]. (**E** and **F**) Mean spatial-overlap matrix (E) and median spatial-distance matrix (F) for the same region derived from multiplexed STORM imaging. Each element of the matrix corresponds to the mean value of the overlap fraction (E) and median value of the center-of-mass distance (F) between a pair of the chromatin segments across ~250 imaged chromosomes. (**G**) Correlation between the Hi-C contact frequency and the mean spatial overlap shown in (D) and (E), respectively. (**H**) Correlation between the Hi-C contact frequencies and median spatial distances shown in (D) and (F), respectively. (**I**) Median spatial-distance matrix for the same genomic region derived from multiplexed diffraction-limited imaging of ~1200 chromosomes. (**J**) Correlation between the Hi-C contact frequencies and median spatial distances shown in (D) and (I), respectively. The Pearson correlation coefficients (ρ) are 0.92, −0.92, and −0.96 in (G), (H), and (J), respectively. The red lines in (H) and (J) are power-law fits with scaling exponents (s) equal to −4.93 ± 0.07 and −4.99 ± 0.05 in (H) and (J), respectively.

structures) similar to those observed in the ensemble Hi-C contact-frequency matrix of the same genomic region (Fig. 1D). Both the spatial overlap and the spatial distance displayed high correlations with the Hi-C contact frequency, with Pearson correlation coefficients of 0.92 and −0.92, respectively (Fig. 1, G and H). Notably, at the kb-to-Mb scale investigated here, the Hi-C contact frequency showed a power-law scaling with the spatial distance, with a scaling exponent of −4.9, similar to our previous observation at the Mb–to–whole chromosome scale (*33*), suggesting that this scaling is potentially a universal property.

We also performed diffraction-limited 3D imaging of each chromatin segment in the same multiplex fashion. The ensemble spatial-distance matrix derived from diffraction-limited imaging also showed similar domain structures (Fig. 1I and fig. S2B) and high correlation with the ensemble Hi-C contact-frequency matrix (Fig. 1J). The multiplexed STORM and diffraction-limited imaging methods have complementary capabilities: The former provided high-resolution information not accessible to the latter, such as the sizes and shapes of individual chromatin segments, whereas the latter allowed for faster image acquisition and higher throughput for the number of chromatin segments or cells imaged. Because the otherwise unresolvable chromatin segments can be separated and localized with high precision by sequential imaging, even the latter approach provided a super-resolution view of chromatin conformation, albeit not as high resolution as that obtained with the STORM images. We used both approaches to interrogate chromatin organization at the single-cell level, as described below.

Super-resolution chromatin tracing reveals TAD-like structures in single cells

Notably, our STORM images and spatial-overlap matrices of individual chromosomes in single cells often showed clear domain structures with higher intradomain chromatin contact (overlap) frequency (Fig. 2, A and B). We refer to these domains as TAD-like structures because of their similar appearance to TADs and sub-TADs in the ensemble-averaged contact matrices, although the boundaries of these single-cell domains varied from cell to cell. For instance, in the STORM images of two example cells, chromatin regions that belong to different ensemble sub-TADs showed extensive overlap in one cell but clear segregation in another cell (Fig. 2, A and B). We identified and quantified the boundary positions of these TAD-like domains from the single-cell spatial-overlap matrices in an automated manner. The domain boundaries showed substantial cell-to-cell variation and a nonzero probability of residing at any of the genomic positions throughout the imaged region (Fig. 2C). Moreover, the domain boundaries exhibited a preference to reside at genomic positions containing strong binding peaks of CTCF and cohesin (marked by one of its core subunits, RAD21), as detected by chromatin immunoprecipitation sequencing (ChIP-seq) (*38*) (Fig. 2, C and D), giving rise to the tendency for ensemble TAD and sub-TAD boundaries to align with these sites (*4*–*6*, *12*). Notably, these domains often appeared as spatially segregated globular structures in the STORM images (Fig. 2B and fig. S3, A and B). We quantified the spatial segregation using a separation score at each genomic

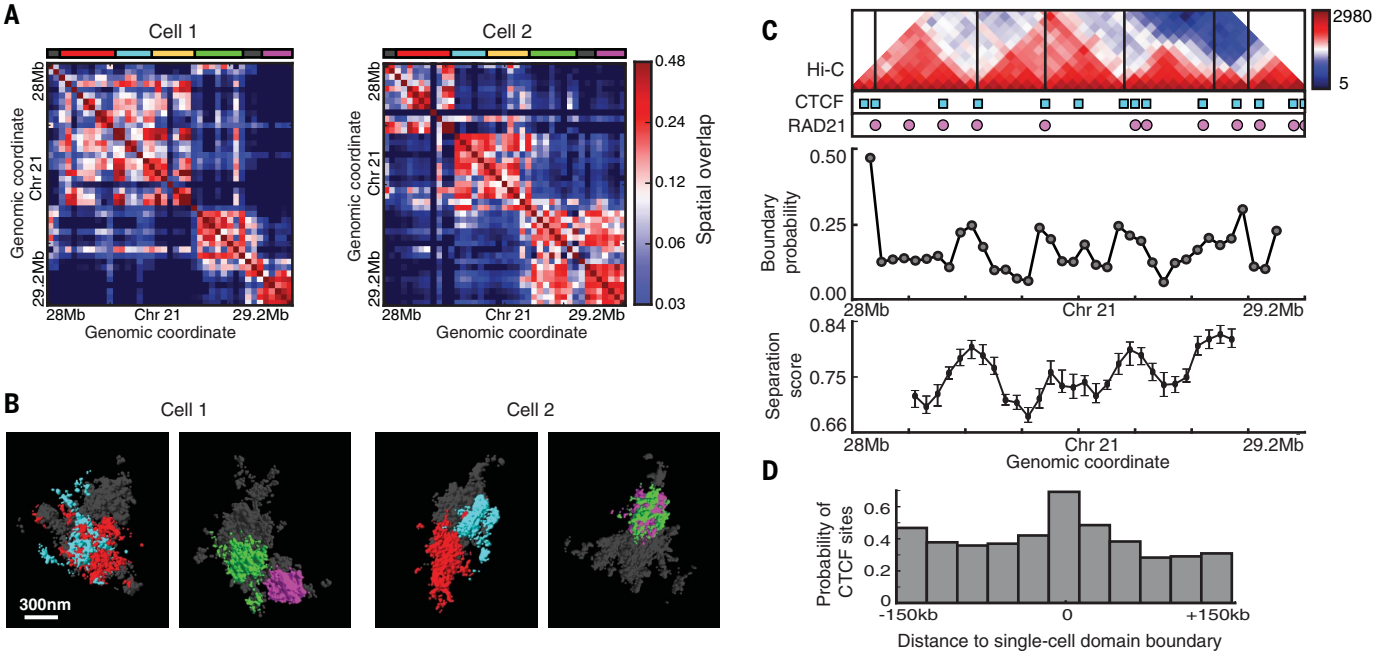


Fig. 2. Chromatin forms TAD-like domain structures with spatially segregated globular conformations in single cells. (**A**) The spatial-overlap matrices of the 1.2-Mb genomic region (Chr21:28Mb-29.2Mb) imaged in one of the two copies of Chr21 from two individual IMR90 cells. The genomic regions marked in red, cyan, yellow, green, and purple correspond to the five sub-TADs observed at the population-average level. (**B**) Multiplexed 3D STORM images corresponding to the two chromosomes shown in (A). The chromatin segments comprising two pairs of ensemble sub-TADs marked as red and cyan or green and purple in (A) are pseudocolored in the same color code. Only one pair of sub-TADs is highlighted in colors per image for ease of visualization, and the other segments in the region of interest are displayed in gray. Each chromatin image is rotated independently to allow the best visualization of the color-highlighted chromatin regions. (**C**) Top: Ensemble Hi-C contact frequency map with sub-TAD boundaries indicated with black lines, shown together with the sites bound by CTCF (cyan squares) and cohesin (represented by RAD21, magenta circles), as determined by ChIP-seq in IMR90 cells (*38*). Middle: The probability (fraction of the ~250 imaged chromosomes) for each genomic location to appear as a single-cell domain boundary. Bottom: The median separation score for each genomic location across the ~250 imaged chromosomes. Error bars indicate 95% confidence intervals derived by resampling (n ~ 250 chromosomes). The separation score is determined as shown in fig. S3. (**D**) The occurrence probability of CTCF and cohesin sites as a function of genomic distance from single-cell domain boundaries. Individual single-cell domain boundaries were aligned, and the relative positions of CTCF ChIP peaks (that colocalize with RAD21 peaks) up to 150 kb on either side of the domain boundaries were histogrammed at 30-kb resolution. The histograms were normalized by dividing by the total number of boundaries.

position describing the level of spatial separation between chromatin on either side of the position (fig. S3, A and B). The separation scores displayed nearly complete segregation at many of the identified single-cell domain boundaries (fig. S3 and Fig. 2C).

Next, we took advantage of the higher throughput of diffraction-limited multiplexed imaging to investigate an extended 2-Mb genomic region (Chr21:28Mb-30Mb) in thousands of individual cells from three distinct cell lines: IMR90 lung fibroblasts, K562 erythroleukemia, and A549 lung epithelial carcinoma cells. Previous ensemble Hi-C data for IMR90 and K562 showed that this genomic region contains two TADs with cell type–specific sub-TADs (fig. S4A) (*12*). The ensemble spatial distance and contact matrices derived from our imaging data again agree well with the ensemble Hi-C contact matrices for IMR90 and K562 and additionally allowed TAD and sub-TAD identifications in A549 cells (Fig. 3, A to C, and fig. S4). At the single-cell level, we observed TAD-like structures with globular 3D conformation and sharp domain boundaries in the images and spatial-distance matrices of individual chromosomes in all three cell types (Fig. 3, D to F). The domain boundary positions again showed cell-to-cell heterogeneity, with a nonzero probability of being located at any genomic positions within the 2-Mb imaged region, and showed a preference for residing at positions bound by CTCF and cohesin (Fig. 3, G to I, and fig. S5).

To test whether these cell-to-cell variations in domain boundaries were caused by different cell-cycle states, we used immunolabeling of the cell-cycle regulator, geminin, along with the 4′,6-diamidino-2-phenylindole (DAPI) stain to separate the cells into G_1, S, and G_2 phases approximately (fig. S6, A and B). The ensemble spatial-distance matrices were similar among all three phases, without any notable change in the TAD and sub-TAD boundaries (fig. S6C), consistent with previous results (*21*, *39*). We observed moderate changes in TAD strength, characterized by the TAD insulation score, suggesting a moderate weakening of ensemble TADs from G_1 to G_2 (fig. S6D), also in agreement with previous results (*21*). Notably, single-cell TAD-like structures were observed in all three phases (fig. S6E), and the cell-to-cell variability and preferential positioning of domain boundaries were all similar among the three phases (fig. S6, F and G), suggesting that the observed cell-to-cell variations in domain boundary positions were not primarily due to differences in cell-cycle state.

Next, we imaged an additional 2-MB region (Chr21:18.6Mb-20.6Mb) that contained no discernable TAD boundaries in the ensemble Hi-C contact matrix (fig. S7A) (*12*). Similarly, our ensemble spatial-distance matrix derived from imaging also did not show any discernable

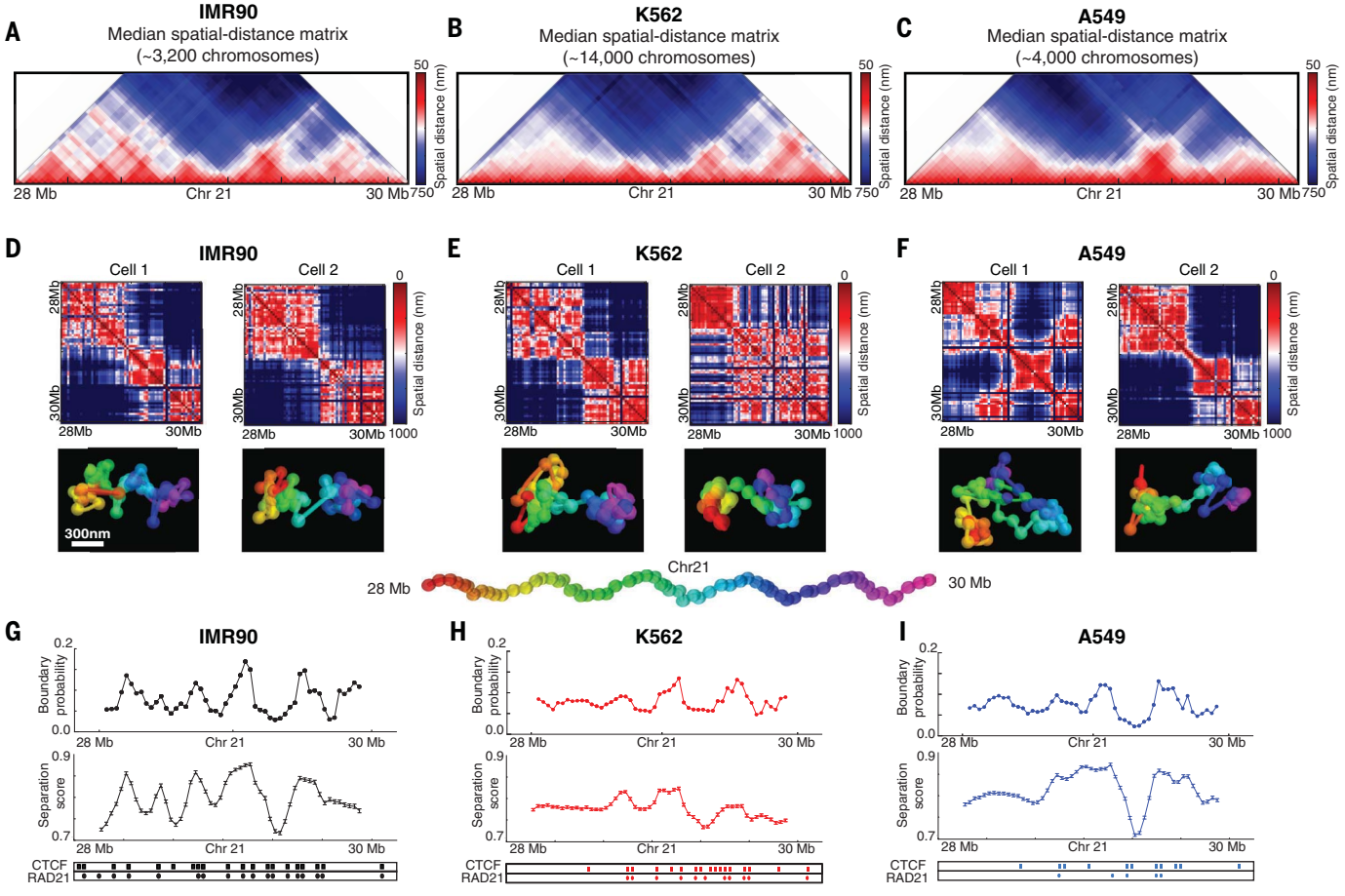


Fig. 3. Single-cell TAD-like structures are formed across cell types. (**A** to **C**) Median spatial-distance matrices for the 2-Mb genomic region of interest (Chr21:28Mb-30Mb) in three cell types: IMR90 lung fibroblast (A), K562 erythroleukemia (B), and A549 carcinomic epithelial cells (C). The number of chromosomes imaged (~3000 to 14,000) is indicated above each matrix. (**D** to **F**) Single-cell spatial-distance matrices of the imaged region (upper) and the corresponding pseudocolored images showing 3D positions of the chromatin segments in each chromosome (lower). Two example cells (and one chromosome copy from each cell) are shown for each of the three cell types [(D) IMR90, (E) K562, (F) A549]. (**G** to **I**) Top: The probability for each genomic position to be a boundary of a single-cell domain for each of the three cell types [(G) IMR90, (H) K562, (I) A549]. Bottom: The mean separation score for each genomic coordinate for each cell type. Error bars indicate 95% confidence intervals (n ~3200, 14,000, and 4000 chromosomes for IMP90, K652, and A549 cells, respectively). The binding sites of CTCF and cohesin (marked by RAD21) determined by ChIP-seq for each cell type (*38*) are indicated with squares and circles, respectively.

domains (fig. S7B). However, the single-cell spatial-distance matrices often showed clearly visible TAD-like structures with sharp domain boundaries (fig. S7C), and the average domain boundary strength was similar to that observed for the ensemble TAD- and subTAD-containing 2-Mb region (Chr21:28Mb-30Mb) described above (fig. S7D). In contrast to the Chr21:28Mb-30Mb region, the Chr21:18.6Mb-20.6Mb region showed largely uniform probability for the presence of single-cell domain boundaries throughout the region (fig. S7E), explaining the lack of domain boundaries in the ensemble matrix of this region.

TAD-like structures remain in single cells after cohesin depletion

We next investigated how chromatin structures change upon removal of the architectural protein, cohesin. DNA extrusion by cohesin complexes has been proposed as a mechanism responsible for ensemble TAD formation (*22*, *40*). Previous studies have shown that the depletion of cohesin causes elimination of TADs at the ensemble level, whereas the A/B compartment structures enriched for active and inactive chromatin are retained (*25*, *41*).

Using HCT116 cells with an auxin-inducible degron fused to a core cohesin subunit RAD21 (*42*), we compared the chromatin structures under both induced and uninduced conditions (Fig. 4 and figs. S8 to S10). We imaged a 2.5-MB region (Chr21:34.6Mb-37.1Mb) that shows several pronounced TAD structures within a single type A compartment (fig. S8A, top panel), in addition to the genomic region described earlier (Chr21:28Mb-30Mb), which shows less pronounced TAD boundaries superimposed on A/B compartment structures in ensemble Hi-C matrices of HCT116 cells (fig. S9, A to C, top panels) (*25*).

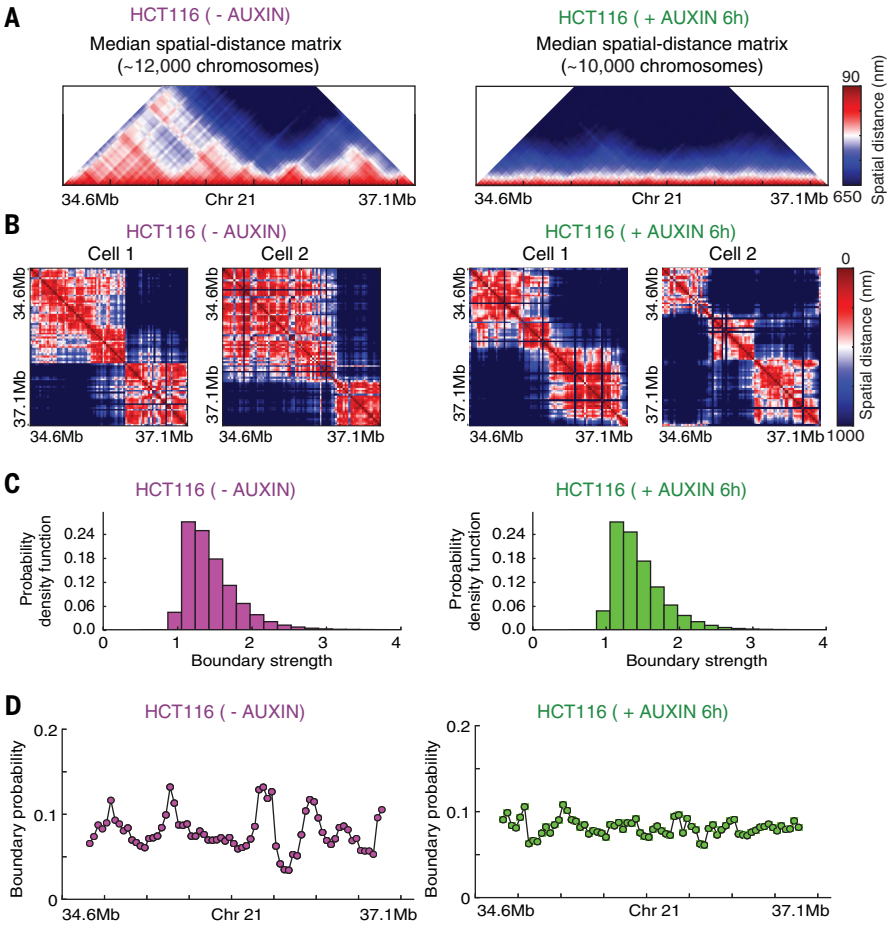


Fig. 4. Single-cell TAD-like structures are present in cells lacking a functional cohesin complex. (**A**) Median spatial-distance matrices for the 2.5-Mb genomic region of interest (Chr21:34.6Mb-37.1Mb) in the transgenic HCT116 cell line without (left) or with (right) auxin treatment to induce cohesin degradation. (**B**) Example single-cell spatial-distance matrices without (left) and with (right) auxin treatment. (**C**) The distribution of boundary strengths in the imaged region for cells without (left) and with (right) auxin treatment. For each identified domain boundary on a single-cell spatial-distance matrix, the boundary strength describes how steeply the spatial distance changed cross the boundary position. The medians of the two distributions with and without auxin treatment differed by less than 1%. (**D**) The probability for each genomic position to be a single-cell domain boundary in cells without (left) or with (right) auxin treatment.

The ensemble spatial-distance matrices derived from our imaging data were similar to the ensemble Hi-C matrices for both regions (Fig. 4A, left panel; fig. S8B, top panel; and fig. S9, D to F, top panels). As expected, upon 6 hours of auxin treatment to induce cohesin degradation, the population-averaged TADs and sub-TADs within the imaged regions were largely eliminated, whereas the A/B compartment structures were retained (Fig. 4A, right panel; fig. S8B, bottom panel; and fig. S9, D to F, bottom panels), consistent with Hi-C results (fig. S8A, bottom panel, and fig. S9, A to C, bottom panels) (*25*). Notably, the chromatin domains observed in single cells persisted after cohesin degradation (Fig. 4B and fig. S10A). Moreover, the domain boundary strengths remained similar between cells with and without cohesin (Fig. 4C and fig. S10B), and the average number of boundaries within the regions also remained similar between cells with and without cohesin, as reflected by the similar values for the mean probability for identifying a domain boundary averaged over all genomic positions (Fig. 4D and fig. S10C). What was notably different in the absence of a functional cohesin complex was that the positions of these domain boundaries became largely uniformly distributed along the genomic coordinate and no longer exhibited preferential positioning at CTCF and cohesin sites as observed in the presence of cohesin (Fig. 4D and fig. S10C). These results indicate that cohesin is not required for the maintenance of TAD-like structures in single cells and that the role of cohesin in the formation of ensemble TADs is to establish preferred genomic boundaries for the single-cell domains. By contrast, preferential boundary positions for A/B compartment were still observed in individual cohesin-depleted cells, similar to those observed for untreated cells (fig. S10D), consistent with the observation that A/B compartments at the ensemble level were retained after cohesin depletion.

We noticed that cohesin depletion strongly hindered but did not completely stop cell division in HCT116 cells (fig. S11, A and B), which allowed us to examine chromatin structures in cells that had gone through a cell cycle without cohesin. We added the modified base 5-ethynyl-2′-deoxyuridine (EdU) to cells for an additional 12 hours in the presence of auxin after 6 hours of initial auxin treatment and used the EdU and geminin signals to select cells that likely had passed through mitosis and reentered G_1 phase (EdU+/geminin− cells) (fig. S11, C to F). We observed that, despite the removal of ensemble TAD boundaries (fig. S11G), single-cell TAD-like domain structures also remained in this population of EdU+/geminin− cells (fig. S11, H to J). Because cells undergo major chromatin reorganization during mitosis (*39*), these TAD-like structures were likely reestablished after mitosis. These data thus suggest that cohesin may not be required for the establishment of domain separation in single cells either, although future experiments are needed to further test this notion.

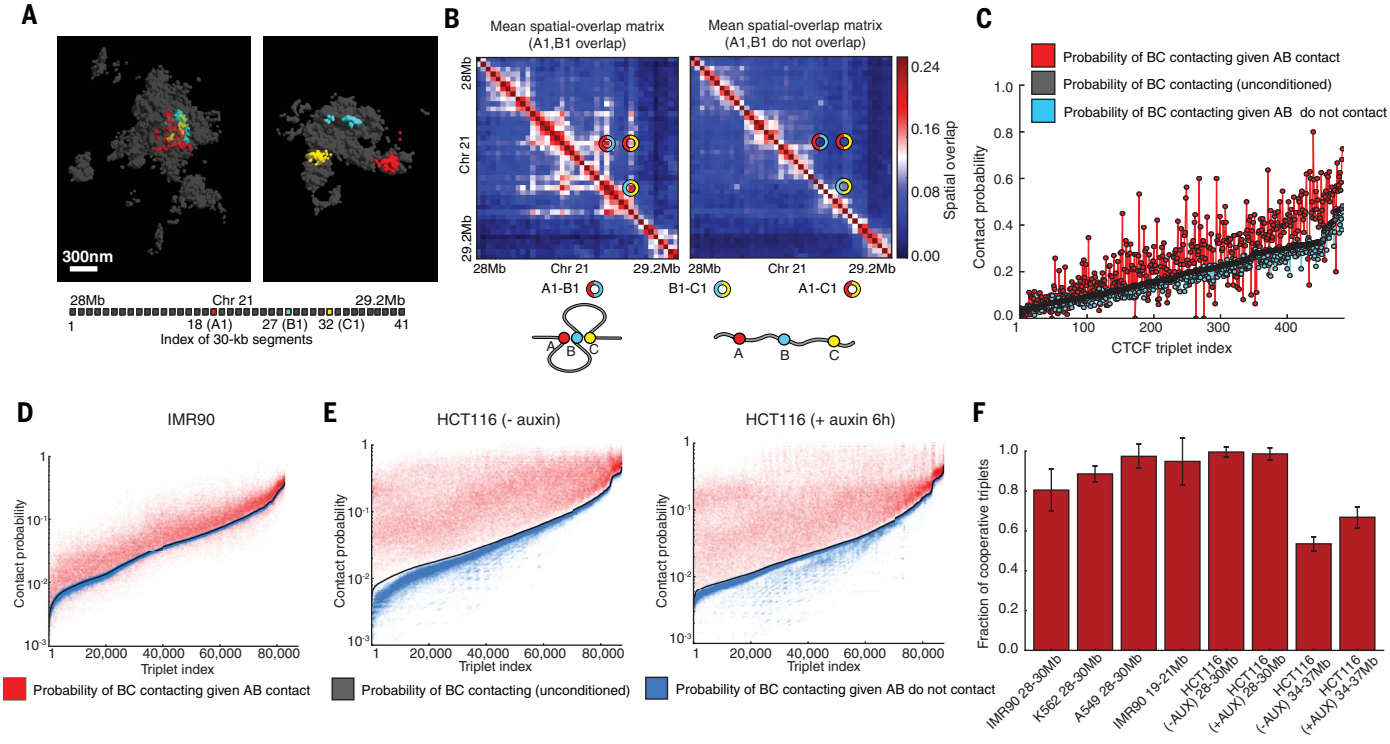


Fig. 5. Cooperative three-way interactions between chromatin segments. (**A**) 3D STORM images of a 1.2-Mb region of interest (Chr21:28Mb-29.2Mb) in one of the two copies of Chr 21 in two different IMR90 cells. The entire genomic region is represented in gray, and three specific 30-kb segments harboring CTCF sites—segments 18 (A1), 27 (B1), and 32 (C1)—are highlighted in red, cyan, and yellow, respectively. (**B**) Cooperative interactions between a specific triplet of segments A1, B1, and C1. Left: The mean spatial-overlap matrix in the subpopulation of chromosomes where segments A1 and B1 overlap. Right: The mean spatial-overlap matrix in the other subpopulation of chromosomes where segments A1 and B1 do not overlap. Circles indicate the matrix elements corresponding to segment pairs A1-B1 (red-cyan), B1-C1 (cyan-yellow), and A1-C1 (red-yellow). (**C**) Cooperative interactions between all possible CTCF-site triplets in the 1.2-Mb imaged region. Shown in the plot are probabilities with which segments B and C contact in individual IMR90 cells under the condition that segments A and B contact (red) or do not contact (blue) for all ordered combinations of CTCF triplets (~500 total) in the imaged region. "Ordered" means that B lies between A and C along the genomic coordinate. Also plotted is the unconditioned probability of B and C contacting regardless of whether A and B contact (black). The index of the triplets is sorted such that the unconditional probability is displayed in ascending order. (**D**) As in (C) but for all ordered triplets of chromatin segments in an extended 2-Mb region of interest (Chr21:28Mb-30Mb) regardless of whether the segment contains CTCF sites. There are ~90,000 such triplets in total, among which only ~2000 are CTCF-site triplets (i.e., all three segments containing CTCF-binding sites). (**E**) As in (D) but for the HCT116 cells without (left) or with (right) auxin treatment. There are ~90,000 such triplets in total, among which only ~700 are CTCF-site triplets. (**F**) The fraction of triplets of segments that show cooperative interactions for each imaged region in various cell types and cohesin depletion conditions.

Cooperative, higher-order chromatin interactions are widespread in single cells

In addition, our chromatin tracing approach allowed us to study higher-order interactions between three or more chromatin loci. We first examined if the interaction between two CTCF sites facilitates or inhibits the interaction with a third. Such higher-order interactions showed cell-to-cell variation, as illustrated in the STORM images (Fig. 5A). When two CTCF sites (represented by letters "A" and "B") showed overlap, we frequently noticed an enhanced overlap of both sites with a third CTCF site, represented by the letter "C", as exemplified by the triplet of CTCF sites (A1, B1, and C1) in Fig. 5B. Because we consider only ordered A, B, and C sites such that C was not between A and B on the genomic coordinate, this facilitation effect cannot be trivially explained by the polymeric nature of chromatin. We systematically quantified this type of facilitated chromatin interactions for all ~500 combinations of such ordered triplets of CTCF sites in the STORM-imaged region of IMR90 cells. Among all triplets analyzed, ~80% showed such facilitated interactions (Fig. 5C).

We next asked whether this cooperative interaction is specific to CTCF sites or is generic to other chromatin loci. We analyzed such three-site interactions for all segments in our imaged regions in all cell types studied (IMR90, K562, A549, and HCT116). We observed that, despite notable quantitative differences observed across different genomic regions and different cell types, contact between two chromatin segments in general tended to increase the probability for these segments to contact a third segment, even when the segments did not harbor CTCF sites (Fig. 5, D to F). Moreover, upon auxin-induced cohesin depletion, this facilitated high-order interaction persisted (Fig. 5, E and F).

Discussion

Our multiplexed, super-resolution imaging method allows the 3D organization of chromatin to be traced with nanometer- and kilobase-scale resolution in thousands of single cells. These imaging data directly revealed diverse chromatin configurations in individual cells, providing insights into the nature of chromatin folding. We observed that chromatin in single cells forms TAD-like domain structures with sharp domain boundaries and that these domain structures often adopt globular conformation with strong physical segregation between neighboring domains. The direct visualization of chromatin conformation, high-detection efficiency of individual genomic loci, and high-density single-cell interaction or distance maps offered by our

imaging approach allowed us to identify these single-cell domain structures that are challenging to detect by previous methods. Hence, our data demonstrate that TAD-like domains are physical structures present in single cells and not an emergent property of population averaging. However, the boundaries positions of these single-cell domains show substantial cell-to-cell variation; therefore, the ensemble TAD boundaries are emergent properties of population averaging due to the preferential positioning of single-cell domain boundaries at sites occupied by CTCF and cohesin. The observed cell-to-cell variability may reflect the dynamic nature of the single-cell domains but is not primarily caused by different cell-cycle states. It is also possible that the epigenetic modification profiles vary from cell to cell, contributing to these observed variations in domain boundary positions. Notably, these single-cell domain structures persist even after depletion of cohesin, a treatment that eliminates TADs and sub-TADs at the population-average level. The loop extrusion model (*22*, *40*), in which cohesin complexes extrude DNA until stopped by a pair of CTCF motifs, has been proposed to explain the formation of TADs and sub-TADs at the population level (*22*, *25*, *40*, *41*). However, in its simplest form, loop extrusion does not lead to strong physical segregation of chromatin domains in single cells (*22*). Our data indicate that cohesin is not required for the maintenance of the observed single-cell TAD-like domain structures and is likely not required for the initial establishment of these structures, either. However, the preferential positioning of the single-cell domain boundaries at CTCF sites was abolished after cohesin depletion, suggesting its dependence on cohesin-CTCF interaction, possibly through loop extrusion, thereby explaining the loss of ensemble TADs upon cohesin depletion.

In addition, we observed higher-order interactions between multiple chromatin loci, and many three-way contacts were observed at higher frequencies than expected from the observed frequency of pairwise interactions, indicative of a form of cooperativity. Such cooperative multiway interaction appears to be a general property of chromatin not limited to specific regulatory elements. It has been suggested that the collision between two loop extruders could facilitate three-way chromatin interactions (*26*, *43*, *44*). Our observation that the cooperative three-way interactions occur even after cohesin depletion indicates that these observed higher-order interactions can arise from a mechanism distinct from the cohesin-based loop extrusion, although our observations do not exclude the possibility that the loop extrusion model could function in parallel to induce higher-order chromatin interactions.

Together, our observations of chromatin organization in single cells expand upon the emerging view that genome packaging is more complex than pairwise interactions (*45*). Our imaging method, which provides a high-resolution physical view of chromatin conformation of targeted genomic regions, can complement sequencing-based genome-wide methods for investigating chromatin organization beyond pairwise interactions. The combination of these methods will help us better understand the complex structural landscape of the genome, tackling problems ranging from interactions among multiple cis-regulatory elements to overall folding conformation of the chromosomes.

Methods summary

Each genomic region of interest was divided into 30-kb segments, and target oligonucleotides for these segments were designed computationally. Target oligonucleotides for each segment were concatenated to a unique readout sequence, along with primer regions for selection and amplification, to constitute the primary probes. The primary probes were synthesized through array-based oligo-pool technology and amplified by polymerase chain reaction and in vitro transcription followed by reverse transcription (*32*). These probes were hybridized to cells adhered to glass coverslips. The samples were mounted in a flow chamber connected to a custom fluidics system for iterative readout probe hybridization and imaged with a custom-assembled microscope (*32*, *33*). Fluorescently labeled oligonucleotides complementary to the readout sequences of each segment, i.e., readout probes, were added by the fluidics system, hybridized for 10 to 30 min, and then rinsed out with a wash buffer. The labeled cells were then imaged by STORM and/or diffraction-limited microscopy. After imaging, the signal of the readout probes was extinguished either by stripping off the probes using DNA strand-displacement or by photobleaching the fluorescence, or both, and readout probes complementary to the next readout sequence(s) [associated with the next chromatin segment(s)] were added. The process was repeated until all chromatin segments were imaged, such that the multiplexed image of the whole genomic region can be constructed with high resolution. Detailed probe design, synthesis, and imaging methods, as well as methods for image analysis; for constructing spatial overlap, distance, and contact matrices; and for single-cell domain analyses are described in the materials and methods section of the supplementary materials.

REFERENCES AND NOTES

1. W. A. Bickmore, The spatial organization of the human genome. *Annu. Rev. Genomics Hum. Genet.* **14**, 67–84 (2013). doi: 10.1146/annurev-genom-091212-153515; pmid: 23875797
2. M. Levine, C. Cattoglio, R. Tjian, Looping back to leap forward: Transcription enters a new era. *Cell* **157**, 13–25 (2014). doi: 10.1016/j.cell.2014.02.009; pmid: 24679523
3. J. Dekker, T. Misteli, Long-Range Chromatin Interactions. *Cold Spring Harb. Perspect. Biol.* **7**, a019356 (2015). doi: 10.1101/cshperspect.a019356; pmid: 26430217
4. J. Dekker, L. Mirny, The 3D Genome as Moderator of Chromosomal Communication. *Cell* **164**, 1110–1121 (2016). doi: 10.1016/j.cell.2016.02.007; pmid: 26967279
5. P. H. L. Krijger, W. de Laat, Regulation of disease-associated gene expression in the 3D genome. *Nat. Rev. Mol. Cell Biol.* **17**, 771–782 (2016). doi: 10.1038/nrm.2016.138; pmid: 27826147
6. M. Yu, B. Ren, The Three-Dimensional Organization of Mammalian Genomes. *Annu. Rev. Cell Dev. Biol.* **33**, 265–289 (2017). doi: 10.1146/annurev-cellbio-100616-060531; pmid: 28783961
7. E. Lieberman-Aiden *et al.*, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009). doi: 10.1126/science.1181369; pmid: 19815776
8. J. R. Dixon *et al.*, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012). doi: 10.1038/nature11082; pmid: 22495300
9. E. P. Nora *et al.*, Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012). doi: 10.1038/nature11049; pmid: 22495304
10. T. Sexton *et al.*, Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458–472 (2012). doi: 10.1016/j.cell.2012.01.010; pmid: 22265598
11. C. Hou, L. Li, Z. S. Qin, V. G. Corces, Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Mol. Cell* **48**, 471–484 (2012). doi: 10.1016/j.molcel.2012.08.031; pmid: 23041285
12. S. S. P. Rao *et al.*, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014). doi: 10.1016/j.cell.2014.11.021; pmid: 25497547
13. J. E. Phillips-Cremins *et al.*, Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013). doi: 10.1016/j.cell.2013.04.053; pmid: 23706625
14. J. H. Gibcus, J. Dekker, The hierarchy of the 3D genome. *Mol. Cell* **49**, 773–782 (2013). doi: 10.1016/j.molcel.2013.02.011; pmid: 23473598
15. T. Sexton, G. Cavalli, The role of chromosome domains in shaping the functional genome. *Cell* **160**, 1049–1059 (2015). doi: 10.1016/j.cell.2015.02.040; pmid: 25768903
16. A. N. Boettiger *et al.*, Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529**, 418–422 (2016). doi: 10.1038/nature16496; pmid: 26760202
17. Q. Szabo *et al.*, TADs are 3D structural units of higher-order chromosome organization in *Drosophila*. *Sci. Adv.* **4**, eaar8082 (2018). doi: 10.1126/sciadv.aar8082; pmid: 29503869
18. T. Nagano *et al.*, Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013). doi: 10.1038/nature12593; pmid: 24067610
19. I. M. Flyamer *et al.*, Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* **544**, 110–114 (2017). doi: 10.1038/nature21711; pmid: 28355183
20. T. J. Stevens *et al.*, 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59–64 (2017). doi: 10.1038/nature21429; pmid: 28289288
21. T. Nagano *et al.*, Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61–67 (2017). doi: 10.1038/nature23001; pmid: 28682332
22. G. Fudenberg *et al.*, Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep.* **15**, 2038–2049 (2016). doi: 10.1016/j.celrep.2016.04.085; pmid: 27210764
23. P. Olivares-Chauvet *et al.*, Capturing pairwise and multi-way chromosomal conformations using chromosomal walks. *Nature* **540**, 296–300 (2016). doi: 10.1038/nature20158; pmid: 27919068
24. R. A. Beagrie *et al.*, Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519–524 (2017). doi: 10.1038/nature21411; pmid: 28273065
25. S. S. P. Rao *et al.*, Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305–320.e24 (2017). doi: 10.1016/j.cell.2017.09.026; pmid: 28985562
26. A. Allahyar *et al.*, Enhancer hubs and loop collisions identified from single-allele topologies. *Nat. Genet.* **50**, 1151–1160 (2018). doi: 10.1038/s41588-018-0161-5; pmid: 29988121
27. S. A. Quinodoz *et al.*, Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* **174**, 744–757.e24 (2018). doi: 10.1016/j.cell.2018.05.024; pmid: 29887377
28. B. Chen *et al.*, Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013). doi: 10.1016/j.cell.2013.12.001; pmid: 24360272
29. M. A. Ricci, C. Manzo, M. F. García-Parajo, M. Lakadamyali, M. P. Cosma, Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell* **160**, 1145–1158 (2015). doi: 10.1016/j.cell.2015.01.054; pmid: 25768910

30. W. Deng, X. Shi, R. Tjian, T. Lionnet, R. H. Singer, CASFISH: CRISPR/Cas9-mediated in situ labeling of genomic loci in fixed cells. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11870–11875 (2015). doi: 10.1073/pnas.1515692112; pmid: 26324940
31. H. Ma *et al*., Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nat. Biotechnol.* **34**, 528–530 (2016). doi: 10.1038/nbt.3526; pmid: 27088723
32. K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, X. Zhuang, Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015). doi: 10.1126/science.aaa6090; pmid: 25858977
33. S. Wang *et al*., Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **353**, 598–602 (2016). doi: 10.1126/science.aaf8084; pmid: 27445307
34. B. J. Beliveau *et al*., Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 21301–21306 (2012). doi: 10.1073/pnas.1213818110; pmid: 23236188
35. B. J. Beliveau *et al*., Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nat. Commun.* **6**, 7147 (2015). doi: 10.1038/ncomms8147; pmid: 25962338
36. M. J. Rust, M. Bates, X. Zhuang, Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**, 793–796 (2006). doi: 10.1038/nmeth929; pmid: 16896339
37. B. Huang, W. Wang, M. Bates, X. Zhuang, Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science* **319**, 810–813 (2008). doi: 10.1126/science.1153529; pmid: 18174397
38. The ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012). doi: 10.1038/nature11247; pmid: 22955616
39. N. Naumova *et al*., Organization of the mitotic chromosome. *Science* **342**, 948–953 (2013). doi: 10.1126/science.1236083; pmid: 24200812
40. A. L. Sanborn *et al*., Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E6456–E6465 (2015). doi: 10.1073/pnas.1518552112; pmid: 26499245
41. W. Schwarzer *et al*., Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017). pmid: 29094699
42. T. Natsume, T. Kiyomitsu, Y. Saga, M. T. Kanemaki, Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. *Cell Reports* **15**, 210–218 (2016). doi: 10.1016/j.celrep.2016.03.001; pmid: 27052166
43. J. H. I. Haarhuis *et al*., The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell* **169**, 693–707.e14 (2017). doi: 10.1016/j.cell.2017.04.013; pmid: 28475897
44. G. A. Busslinger *et al*., Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* **544**, 503–507 (2017). doi: 10.1038/nature22063; pmid: 28424523
45. D. Hnisz, K. Shrinivas, R. A. Young, A. K. Chakraborty, P. A. Sharp, A Phase Separation Model for Transcriptional Control. *Cell* **169**, 13–23 (2017). doi: 10.1016/j.cell.2017.02.007; pmid: 28340338

ACKNOWLEDGMENTS

We thank N. Kleckner and M. Kanemaki for providing the transgenic HCT-116 RAD21-mAID-mClover cell line. **Funding:** This work is supported in part by the National Institutes of Health (to X.Z.) and a Burroughs Wellcome CASI grant and Searle Scholar's grant (to A.N.B.). L.J.M. is supported in part by the Stanford Developmental Biology Training program. J.-H.S. is supported in part by the Harvard MCO training program. N.A.S.-A. was supported by a National Defense Science and Engineering Graduate Fellowship (NDSEG) Program and by a Stanford Graduate Fellowship. M.P. was supported in part by an award from the Stanford Vice President for Undergraduate Research office. S.K. is supported in part by the Harvard Biophysics training program. K.Y. is supported by a Stanford Graduate Fellowship. X.Z. is a Howard Hughes Medical Institute Investigator. **Author contributions:** B.B., A.N.B., and X.Z. designed the experiments. B.B., L.J.M., J.-H.S., N.A.S.-A., M.P., S.K., and K.Y. performed the experiments. B.B., L.J.M., and A.N.B. performed analysis; and B.B., A.N.B., and X.Z. interpreted the results. B.B., A.N.B., and X.Z. wrote the manuscript with input from L.J.M., J.-H.S., N.A.S.-A., M.P., S.K., and K.Y. **Competing interests:** The authors have no competing interests. **Data and materials availability:** All codes for imaging data analyses have been submitted to GitHub and are publicly available (https://github.com/BogdanBintu/ChromatinImaging).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6413/eaau1783/suppl/DC1
Materials and Methods
References (*46*–*56*)
Figs. S1 to S11
Tables S1 and S2

15 May 2018; accepted 4 September 2018
10.1126/science.aau1783

RESEARCH ARTICLE SUMMARY

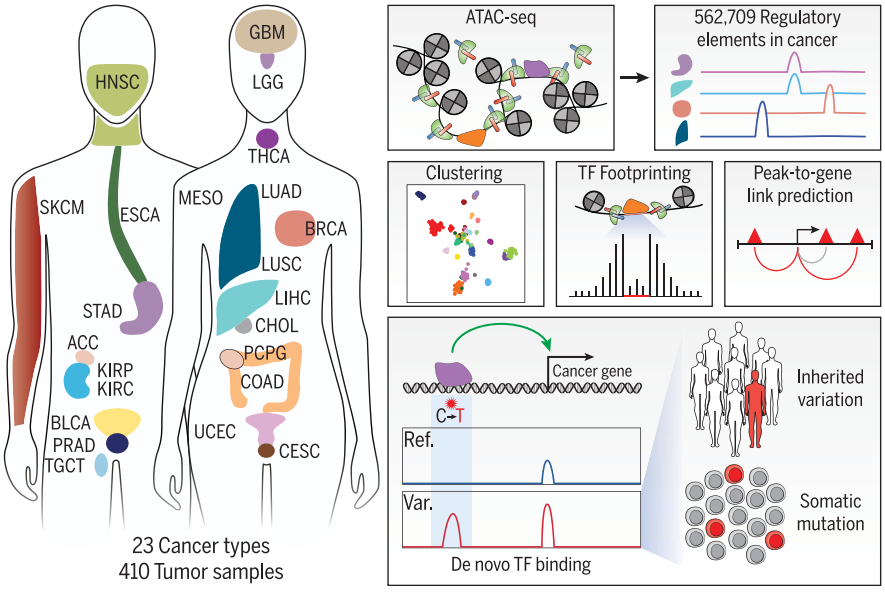
CANCER

The chromatin accessibility landscape of primary human cancers

M. Ryan Corces*, Jeffrey M. Granja*, Shadi Shams, Bryan H. Louie, Jose A. Seoane, Wanding Zhou, Tiago C. Silva, Clarice Groeneveld, Christopher K. Wong, Seung Woo Cho, Ansuman T. Satpathy, Maxwell R. Mumbach, Katherine A. Hoadley, A. Gordon Robertson, Nathan C. Sheffield, Ina Felau, Mauro A. A. Castro, Benjamin P. Berman, Louis M. Staudt, Jean C. Zenklusen, Peter W. Laird, Christina Curtis, The Cancer Genome Atlas Analysis Network, William J. Greenleaf†, Howard Y. Chang†

INTRODUCTION: Cancer is one of the leading causes of death worldwide. Although the 2% of the human genome that encodes proteins has been extensively studied, much remains to be learned about the noncoding genome and gene regulation in cancer. Genes are turned on and off in the proper cell types and cell states by transcription factor (TF) proteins acting on DNA regulatory elements that are scattered over the vast noncoding genome and exert long-range influences. The Cancer Genome Atlas (TCGA) is a global consortium that aims to accelerate the understanding of the molecular basis of cancer. TCGA has systematically collected DNA mutation, methylation, RNA expression, and other comprehensive datasets from primary human cancer tissue. TCGA has served as an invaluable resource for the identification of genomic aberrations, altered transcriptional networks, and cancer subtypes. Nonetheless, the gene regulatory landscapes of these tumors have largely been inferred through indirect means.

RATIONALE: A hallmark of active DNA regulatory elements is chromatin accessibility. Eukaryotic genomes are compacted in chromatin, a complex of DNA and proteins, and only the active regulatory elements are accessible by the cell's machinery such as TFs. The assay for transposase-accessible chromatin using sequencing (ATAC-seq) quantifies DNA accessibility through the use of transposase enzymes that insert sequencing adapters at these accessible chromatin sites. ATAC-seq enables the genome-wide profiling of TF binding events that orchestrate gene expression programs and give a cell its identity.



Cancer gene regulatory landscape. Chromatin accessibility profiling of 23 human cancer types (left) in 410 tumor samples from TCGA revealed 562,709 DNA regulatory elements. The activity of these DNA elements organized cancer subtypes, identified TF proteins and regulatory elements controlling cancer gene expression, and suggested molecular mechanisms for cancer-associated inherited variants and somatic mutations in the noncoding genome. See main article for abbreviations of cancer types. Ref., reference; Var., variant.

RESULTS: We generated high-quality ATAC-seq data in 410 tumor samples from TCGA, identifying diverse regulatory landscapes across 23 cancer types. These chromatin accessibility profiles identify cancer- and tissue-specific DNA regulatory elements that enable classification of tumor subtypes with newly recognized prognostic importance. We identify distinct TF activities in cancer based on differences in the inferred patterns of TF-DNA interaction and gene expression. Genome-wide correlation of gene expression and chromatin accessibility predicts tens of thousands of putative interactions between distal regulatory elements and gene promoters, including key oncogenes and targets in cancer immunotherapy, such as *MYC*, *SRC*, *BCL2*, and *PDL1*. Moreover, these regulatory interactions inform known genetic risk loci linked to cancer predisposition, nominating biochemical mechanisms and target genes for many cancer-linked genetic variants. Lastly, integration with mutation profiling by whole-genome sequencing identifies cancer-relevant noncoding mutations that are associated with altered gene expression. A single-base mutation located 12 kilobases upstream of the *FGD4* gene, a regulator of the actin cytoskeleton, generates a putative de novo binding site for an NKX TF and is associated with an increase in chromatin accessibility and a concomitant increase in *FGD4* gene expression.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aav1898

CONCLUSION: The accessible genome of primary human cancers provides a wealth of information on the susceptibility, mechanisms, prognosis, and potential therapeutic strategies of diverse cancer types. Prediction of interactions between DNA regulatory elements and gene promoters sets the stage for future integrative gene regulatory network analyses. The discovery of hundreds of noncoding somatic mutations that exhibit allele-specific regulatory effects suggests a pervasive mechanism for cancer cells to manipulate gene expression and increase cellular fitness. These data may serve as a foundational resource for the cancer research community. ■

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: howchang@stanford.edu (H.Y.C.); wjg@stanford.edu (W.J.G.)
Cite this article as M. R. Corces *et al.*, *Science* **362**, eaav1898 (2018). DOI: 10.1126/science.aav1898

RESEARCH ARTICLE

CANCER

The chromatin accessibility landscape of primary human cancers

M. Ryan Corces[1]*, Jeffrey M. Granja[1,2,3]*, Shadi Shams[1], Bryan H. Louie[1], Jose A. Seoane[2,4,5], Wanding Zhou[6], Tiago C. Silva[7,8], Clarice Groeneveld[9], Christopher K. Wong[10], Seung Woo Cho[1], Ansuman T. Satpathy[1], Maxwell R. Mumbach[1,2], Katherine A. Hoadley[11], A. Gordon Robertson[12], Nathan C. Sheffield[13], Ina Felau[14], Mauro A. A. Castro[9], Benjamin P. Berman[7], Louis M. Staudt[14], Jean C. Zenklusen[14], Peter W. Laird[6], Christina Curtis[2,4,5], The Cancer Genome Atlas Analysis Network†, William J. Greenleaf[1,2,3,15,16]‡, Howard Y. Chang[1,2,17,18]‡

We present the genome-wide chromatin accessibility profiles of 410 tumor samples spanning 23 cancer types from The Cancer Genome Atlas (TCGA). We identify 562,709 transposase-accessible DNA elements that substantially extend the compendium of known cis-regulatory elements. Integration of ATAC-seq (the assay for transposase-accessible chromatin using sequencing) with TCGA multi-omic data identifies a large number of putative distal enhancers that distinguish molecular subtypes of cancers, uncovers specific driving transcription factors via protein-DNA footprints, and nominates long-range gene-regulatory interactions in cancer. These data reveal genetic risk loci of cancer predisposition as active DNA regulatory elements in cancer, identify gene-regulatory interactions underlying cancer immune evasion, and pinpoint noncoding mutations that drive enhancer activation and may affect patient survival. These results suggest a systematic approach to understanding the noncoding genome in cancer to advance diagnosis and therapy.

Cancer is a highly heterogeneous group of diseases, with each tumor type exhibiting distinct clinical features, patient outcomes, and therapeutic responses. The Cancer Genome Atlas (TCGA) was established to characterize this heterogeneity and understand the molecular underpinnings of cancer (*1*). Through large-scale genomic and molecular analyses, TCGA has revealed an exquisite diversity of genomic aberrations, altered transcriptional networks, and tumor subtypes that have engendered a more comprehensive understanding of disease etiologies and laid the foundations for new therapeutics and impactful clinical trials.

Work from TCGA and many others has demonstrated the importance of the epigenome to cancer initiation and progression (*2*). Profiling of cancer-specific coding mutations through whole-exome sequencing has identified prominent driver mutations in genes encoding chromatin remodeling enzymes and modifiers of DNA methylation. These mutations drive alterations in the epigenome which, in turn, can establish the dysregulated cellular phenotypes that have become known as the hallmarks of cancer (*3*). Although many principles of chromatin regulation have been elucidated in cultured cancer cells, epigenomic studies of primary tumors are especially valuable, capturing the genuine ecosystem of heterotypic tumor and stromal cell interactions and the impacts of factors in the tumor microenvironment such as hypoxia, acidosis, and matrix stiffness (*4*). TCGA has carried out targeted DNA methylation profiling of more than 10,000 samples and, more recently, whole-genome bisulfite sequencing (WGBS) of 39 TCGA tumor samples (*5*). This data-rich resource has identified cancer-specific differentially methylated regions, providing an unprecedented view of epigenetic heterogeneity in cancer. Integration of DNA methylation and additional TCGA data types has enabled the prediction of functional regulatory elements (*6*–*8*) and the identification of previously unknown cancer subtypes (*9*–*13*). Additional work has identified cancer-relevant variable enhancer loci by using histone modifications (*14*) and enhancer RNA sequencing (*15*). These studies represent, to date, the largest genome-wide epigenomic profiling efforts in primary human cancer samples.

Recently, the advent of the assay for transposase-accessible chromatin using sequencing (ATAC-seq) (*16*) has enabled the genome-wide profiling of chromatin accessibility in small quantities of frozen tissue (*17*). Because accessible chromatin is a hallmark of active DNA regulatory elements, ATAC-seq makes it possible to assess the gene regulatory landscape in primary human cancers. Combined with the richness of diverse, orthogonal data types in TCGA, the chromatin accessibility landscape in cancer provides a key link between inherited and somatic mutations, DNA methylation, long-range gene regulation, and, ultimately, gene expression changes that affect cancer prognosis and therapy.

Results

ATAC-seq in frozen human cancer samples is highly robust

We profiled the chromatin accessibility landscape for 23 types of primary human cancers, represented by 410 tumor samples derived from 404 donors from TCGA (protocol S1). These 23 cancer types are representative of the diversity of human cancers (Fig. 1A and data S1). From the 410 tumor samples, we generated technical replicates from 386 samples, yielding 796 genome-wide chromatin accessibility profiles (data S1). Given the size of this cohort, we first ensured that all generated ATAC-seq data could be uniquely mapped to the expected donor through comparison with single-nucleotide polymorphism (SNP) genotyping calls (fig. S1A). In all samples, the genotype from the ATAC-seq data generated in this study correlated most highly with previously published genotyping array data for the expected donor compared with that of all other 11,126 TCGA donors. All ATAC-seq data included in this study passed a minimum threshold of enrichment of signal over background (fig. S1, B to D, and data S1) with most samples showing a characteristic fragment size distribution with clear nucleosomal periodicity (fig. S1E). With this high-quality set of 410 tumor samples, we identified 562,709 reproducible (observed in more than one replicate) pan-cancer

[1]Center for Personal Dynamic Regulomes, Stanford University, Stanford, CA 94305, USA. [2]Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA. [3]Program in Biophysics, Stanford University School of Medicine, Stanford, CA 94305, USA. [4]Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA. [5]Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA 94305, USA. [6]Center for Epigenetics, Van Andel Research Institute, Grand Rapids, MI 49503, USA. [7]Center for Bioinformatics and Functional Genomics, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA. [8]Department of Genetics, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, SP CEP 14.040-905, Brazil. [9]Bioinformatics and Systems Biology Laboratory, Polytechnic Center, Federal University of Paraná, Curitiba, PR CEP 80.060-000, Brazil. [10]Department of Biomolecular Engineering, Center for Biomolecular Sciences and Engineering, University of California–Santa Cruz, Santa Cruz, CA 95064, USA. [11]Department of Genetics, Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. [12]Canada's Michael Smith Genome Sciences Center, BC Cancer Agency, Vancouver, BC V5Z 4S6, Canada. [13]Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA. [14]National Cancer Institute, NIH, Bethesda, MD 20892, USA. [15]Department of Applied Physics, Stanford University, Stanford, CA 94025, USA. [16]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA. [17]Program in Epithelial Biology, Stanford University, Stanford, CA 94305, USA. [18]Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA.

*These authors contributed equally to this work. †The Cancer Genome Atlas Analysis Network collaborators and affiliations are listed in the supplementary materials.

‡Corresponding author. Email: howchang@stanford.edu (H.Y.C.); wjg@stanford.edu (W.J.G.)

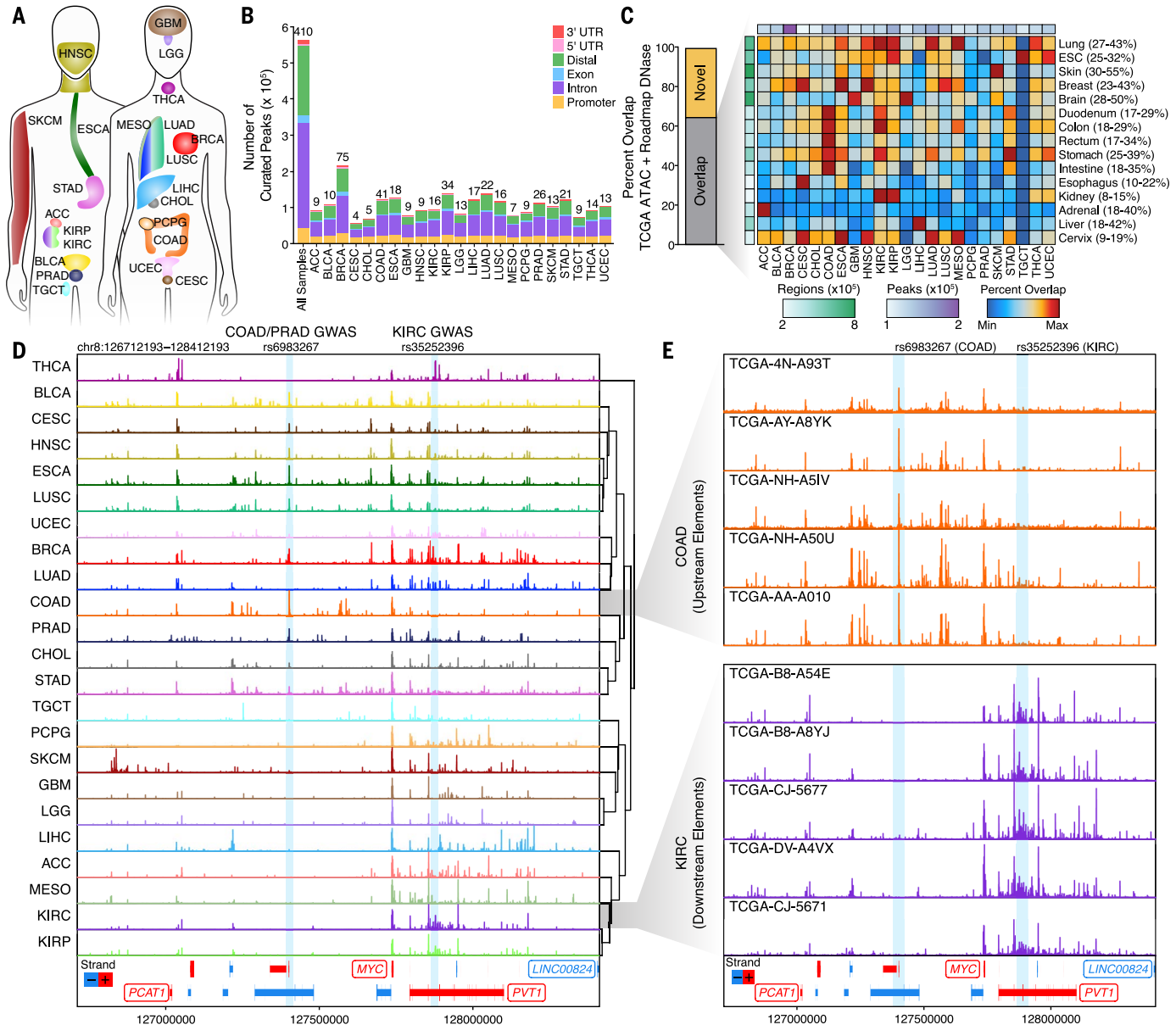


Fig. 1. Pan-cancer ATAC-seq of TCGA samples identifies diverse regulatory landscapes. (**A**) Diagram of the 23 cancer types profiled in this study. Colors are kept consistent throughout figures. (**B**) Pan-cancer peak calls from ATAC-seq data. Peak calls from each cancer type are shown individually in addition to the 562,709 peaks that represent the pan-cancer merged peak set. Color indicates the type of genomic region overlapped by the peak. The numbers shown above each bar represent the number of samples profiled for each cancer type. UTR, untranslated region. (**C**) Overlap of cancer type-specific ATAC-seq peaks with Roadmap DNase-seq peaks from various tissues and cell types. Left: The percent of ATAC-seq peaks that are overlapped by one or more Roadmap peaks. Right: A heatmap of the percent overlap observed for each ATAC-seq peak set within the Roadmap DNase-seq peak set. Colors are scaled according to the minimum and maximum overlaps, which are indicated numerically to the right of the DNase-seq peak set names. The total number of ATAC-seq peaks (white to purple) or Roadmap DNase-seq regions (white to green) are shown colorimetrically. (**D**) Normalized ATAC-seq sequencing tracks of all 23 cancer types at the *MYC* locus. Each track represents the average accessibility per 100-bp bin across all replicates. Known GWAS SNPs rs6983267 (COAD, PRAD) and rs35252396 (KIRC) are highlighted with light blue shading. Region shown represents chromosome 8 (chr8):126712193 to 128412193. (**E**) Normalized ATAC-seq sequencing tracks of five different COAD samples (top, orange) and KIRC samples (bottom, purple) shown across the same *MYC* locus as in Fig. 1D. Known GWAS SNPs rs6983267 (COAD, PRAD) and rs35252396 (KIRC) are highlighted with light blue shading. Region shown represents chr8:126712193 to 128412193. ACC, adrenocortical carcinoma; BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma; CHOL, cholangiocarcinoma; COAD, colon adenocarcinoma; ESCA, esophageal carcinoma; GBM, glioblastoma multiforme; HNSC, head and neck squamous cell carcinoma; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LGG, low grade glioma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; MESO, mesothelioma; PCPG, pheochromocytoma and paraganglioma; PRAD, prostate adenocarcinoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; TGCT, testicular germ cell tumors; THCA, thyroid carcinoma; UCEC, uterine corpus endometrial carcinoma.

peaks of chromatin accessibility (Fig. 1B and data S2). These peaks were identified using a normalized peak score metric to enable direct comparison of peaks across samples of unequal sequencing depth, with each cancer type having an average of 105,585 peaks (range 56,125 to 215,978; Fig. 1B and fig. S1F; see methods). Reproducibility within the pan-cancer peak set was high for technical replicates (different nuclei from the same tumor sample; fig. S1, G and H), intratumor replicates (different samples from the same tumor; fig. S1I), and intertumor replicates (tumor samples from different donors; fig. S1, J and K).

Cancer chromatin accessibility extends the dictionary of DNA regulatory elements

The pan-cancer and cancer type–specific peak sets generated in this study enabled quantification of the number of DNA regulatory elements identified. To do this, we compared the regions defined by our pan-cancer and cancer type–specific peak sets to the regions defined by the Roadmap Epigenomics Project deoxyribonuclease I hypersensitive sites sequencing (DNase-seq) studies (*18*), finding a median of 34.4% overlap between the cancer type–specific peak sets and the various Roadmap tissue-type peak sets, with the strongest overlap occurring in the expected combinations (Fig. 1C and data S3). In total, about 65% of the pan-cancer peaks identified in this study had overlap with previously observed regulatory elements, highlighting both the consistency of our results with published datasets and the large number of additional putative regulatory elements observed in this study (Fig. 1C). Given the extensive coverage of Roadmap DNase-seq studies in healthy tissues, our results suggested that the disease context of cancer unveils the activity of additional DNA regulatory elements. Moreover, overlap of the ATAC-seq–defined DNA regulatory elements with chromatin immunoprecipitation sequencing (ChIP-seq)–defined ChromHMM regulatory states shows a strong enrichment of accessible chromatin sites in promoter and enhancer regions, as expected (fig. S1L). Although we profiled many samples in some cancer types [i.e., breast invasive carcinoma (BRCA), 75 tumor samples], we profiled fewer samples in multiple other cancer types (i.e., cervical squamous cell carcinoma, four tumor samples) (Fig. 1B). By estimating the number of unique peaks added with each additional sample, we found that cancer types have an estimated average of 169,822 total peaks (range 97,995 to 309,313) at saturation (fig. S1, M and N, and data S3), suggesting that profiling of additional samples of each cancer type would further expand the repertoire of regulatory elements.

Noncoding DNA elements reveal distinct cancer gene regulation and genetic risks

The *MYC* proto-oncogene locus provides a prime illustration of the diversity of the chromatin accessibility landscape across cancer types. *MYC* is embedded in a region with multiple DNA regulatory elements and noncoding transcripts that regulate *MYC* in a tissue-specific fashion (*19*). We observed sufficient diversity in the chromatin accessibility landscape of the *MYC* locus to enable clustering of cancer types into two primary categories: (i) cancer types with extensive chromatin accessibility at 5′ and 3′ DNA elements, such as colon adenocarcinoma (COAD), and (ii) cancer types with chromatin accessibility primarily at 3′ regulatory elements, such as kidney renal clear cell carcinoma (KIRC) (Fig. 1D). This trend is consistent across different samples of the same cancer type, as shown for COAD and KIRC (Fig. 1E) and is similar to the regulation observed in the *HOXD* locus (*20*).

Genome-wide association studies (GWAS) have identified numerous inherited risk loci for cancer susceptibility. However, many of these SNPs reside in the noncoding genome within known DNA regulatory elements. In the *MYC* locus, we identify known sites of chromatin accessibility, including peaks surrounding functionally validated GWAS cancer susceptibility SNPs (rs6983267 and rs35252396; Fig. 1, D and E). SNP rs6983267 is associated with increased susceptibility to colon adenocarcinoma and prostate adenocarcinoma (PRAD) (*21*–*23*), consistent with the presence of focal chromatin accessibility in these cancer types. However, SNP rs6983267 has not been previously associated with breast cancer or any squamous tumor types, which also have strong chromatin accessibility at this regulatory element in our ATAC-seq data (Fig. 1D). Similarly, SNP rs35252396 has been associated with KIRC and, in our data, shows strong accessibility in samples from kidney cancer types as well as breast and thyroid carcinoma, suggesting a potential role for these SNPs in previously unappreciated cancer contexts.

To visualize global patterns from our diverse ATAC-seq datasets, we performed Pearson correlation hierarchal clustering on distal and promoter elements (Fig. 2A). We found that distal elements exhibited a greater specificity and wider dynamic range of activity in association with cancer types, whereas promoter element accessibility was less cancer type–specific and showed similar patterns of correlation to global gene expression, as measured by RNA-seq (Fig. 2A). This functional specificity of distal regulatory elements was also previously observed in healthy tissues and in development (*24*, *25*). Using *t*-distributed stochastic neighbor embedding (*26*) (t-SNE; Fig. 2B) and density clustering (*27*) (fig. S2A), we identified 18 distinct clusters, which we labeled based on the observed cancer-type enrichment (fig. S2B and data S3). We found strong concordance between this ATAC-seq–based clustering and the published multiomic iCluster scheme using TCGA mRNA-seq, microRNA (miRNA)–seq, DNA methylation, reverse-phase protein array (RPPA), and DNA copy number data (*28*) (Fig. 2, C and D). Comparing this clustering scheme to other TCGA-based clustering schemes, we observed the strongest concordance of our ATAC-seq clustering scheme with mRNA and cancer type (Fig. 2E). This is consistent with the connection of chromatin accessibility to transcriptional output and the observation that ATAC-seq is strongly cell type–specific. Multiple observations can be made from these clusters: (i) Some cancer types split into two distinct clusters such as breast cancer (i.e., basal and nonbasal) and esophageal cancer (i.e., squamous and adenocarcinoma), (ii) cancer samples derived from the same tissue type often group together [i.e., kidney renal papillary cell carcinoma (KIRP) and KIRC], and (iii) some cancers group together across tissues as observed for squamous cell types (Fig. 3A and fig. S2B).

Cluster-specific regulatory landscapes identify patterns of transcription factor usage and DNA hypomethylation

Grouping of samples into defined clusters enables the determination of patterns in chromatin accessibility that are unique to each cluster. Using a framework that we term "distal binarization," we identified the distal regulatory elements that are accessible only in a single cluster or small group of clusters (Fig. 3B, fig. S2C, and data S4). Of the 516,927 pan-cancer distal elements, 203,260 were found to be highly accessible in a single cluster or group of clusters (up to four clusters). These cluster-specific peak sets are enriched for motifs of transcription factors (TFs) with correlated gene expression that are known to be important for cancer and tissue identity (Fig. 3C, fig. S2D, and data S4). These include the androgen receptor (AR) in prostate cancer, forkhead box A1 (FOXA1) in nonbasal breast cancer, and melanogenesis-associated transcription factor (MITF) in melanoma. Moreover, these cluster-specific peak sets are enriched for known GWAS SNPs that are associated with cancers of the corresponding type (fig. S2E and data S5), highlighting that cancer-related GWAS SNPs tend to be located within or near cancer type–specific regulatory elements. The concordance of GWAS risk loci and cancer chromatin state has often been evaluated using cancer cell lines in the past, and our work provides a foundational map to evaluate noncoding GWAS SNPs in primary human cancers.

Consistent with published reports (*12*, *18*, *29*, *30*), the degree of DNA methylation was anticorrelated with chromatin accessibility at regulatory elements, and regions lacking chromatin accessibility were more frequently methylated (fig. S2F). In particular, cluster-specific peak sets are hypomethylated in the relevant cancer types, though frequently methylated in other cancer types that lack accessibility in those peaks (fig. S2G). Consistent with these observations, which are based on DNA methylation array data, we see a strong depletion of DNA methylation at the center of both distal peaks and promoter peaks in a single patient profiled by WGBS (fig. S2H) (*5*). In our analysis of methylation levels within cluster-specific peak sets, we also identified a subgroup of brain cancers that exhibits DNA hypermethylation of peaks specific to nonbrain cancers (fig. S2G), likely caused by mutations in genes

that affect DNA methylation, such as isocitrate dehydrogenase 1 (*IDH1*) (fig. S3A). Similarly, we found that the subset of testicular germ cell tumors that are seminomas show a pattern of genome-wide DNA hypomethylation, consistent with a published report (*31*) (fig. S3B). Thus, a small number of TFs dominate the cis-regulatory landscape in each cancer type. These TFs are often the known key drivers of the respective cancer or tissue type, and TF occupancy is associated with, and possibly causes, DNA hypomethylation of the corresponding DNA elements in cancer.

De novo identification of cancer subtypes from ATAC-seq data

Given the richness of the chromatin accessibility landscape, we explored the capacity of ATAC-seq data to define molecular subtypes of cancer de novo. This analysis was limited to cancer types with sufficient available donors: BRCA (N = 74), PRAD (N = 26), and KIRP (N = 34). In KIRP, a gap statistic identified three distinct subgroups that are clearly separable by the first two principal components (Fig. 3D). The smallest of these subgroups contains four donors with very clear differences in ATAC-seq

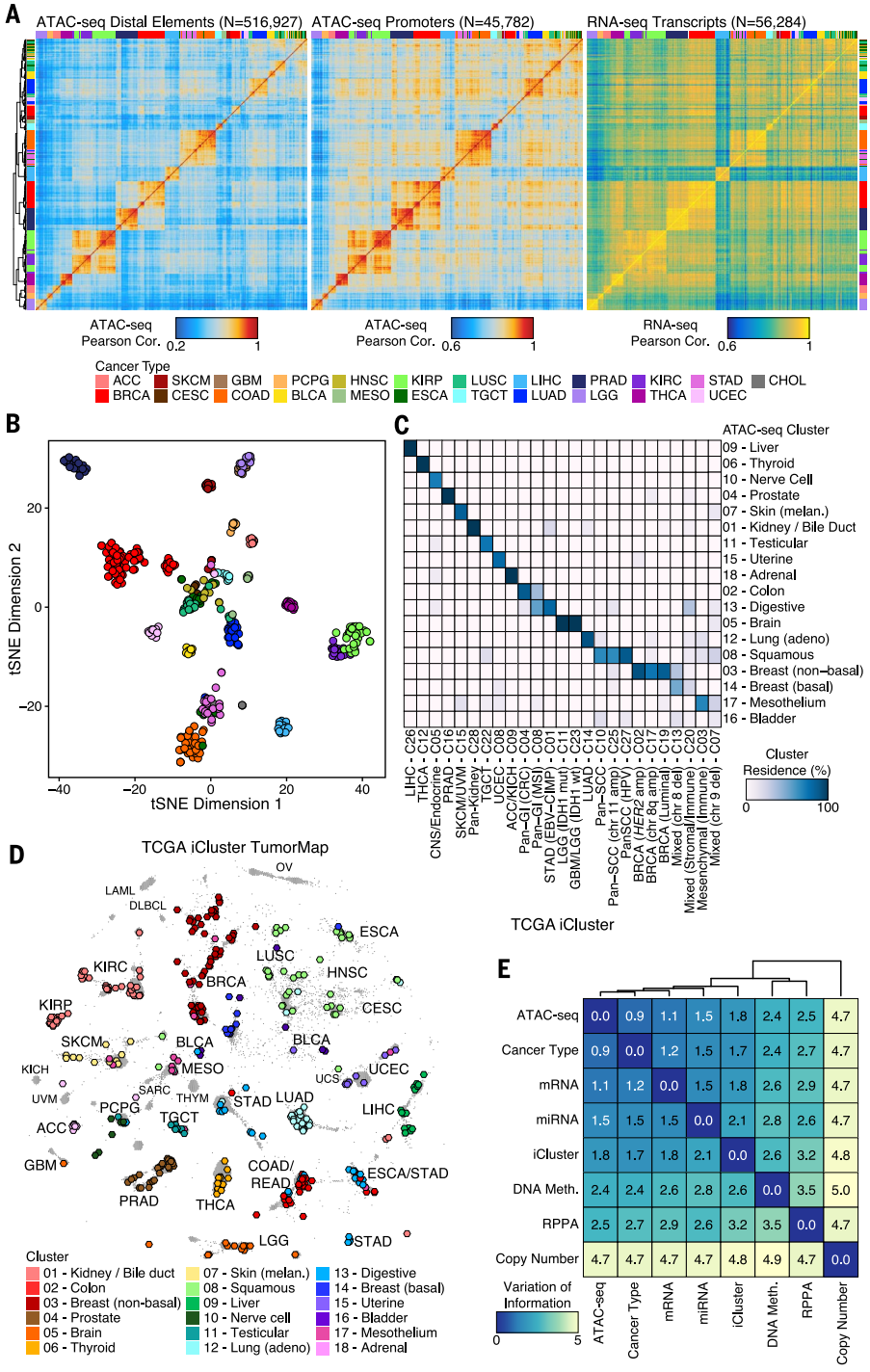


Fig. 2. Chromatin accessibility profiles reveal distinct molecular subtypes of cancers. (**A**) Pearson correlation heatmaps of ATAC-seq distal elements (left), ATAC-seq promoters (middle), and RNA-seq of all genes (right). Clustering orientation is dictated by the ATAC-seq distal element accessibility, and all other heatmaps use this same clustering orientation. Color scale values vary between heatmaps. Promoter peaks are defined as occurring between −1000 and +100 bp of a transcriptional start site. Distal peaks are all nonpromoter peaks. The total number of features (N) used for correlation is indicated above each Pearson correlation heatmap. (**B**) Unsupervised t-SNE on the top 50 principal components for the 250,000 most variable peaks across all cancer types. Each dot represents the merge of all technical replicates from a given sample. Color represents the cancer type shown above the plot. (**C**) Cluster residence heatmap showing the percent of each TCGA iCluster that overlaps with each ATAC-seq–based cluster. (**D**) ATAC-seq t-SNE clusters shown on the PanCanAtlas iCluster TumorMap. Each hexagon represents a cancer patient sample, and the positions of the hexagons are computed from the similarity of samples in the iCluster latent space. The color and larger size of the hexagon indicates the ATAC-seq cluster assignment. Samples that were not included in the ATAC-seq analysis are represented by smaller gray hexagons. The text labels indicate the cancer disease type. (**E**) Variation of information analysis of clustering schemes derived by using various data types from TCGA.

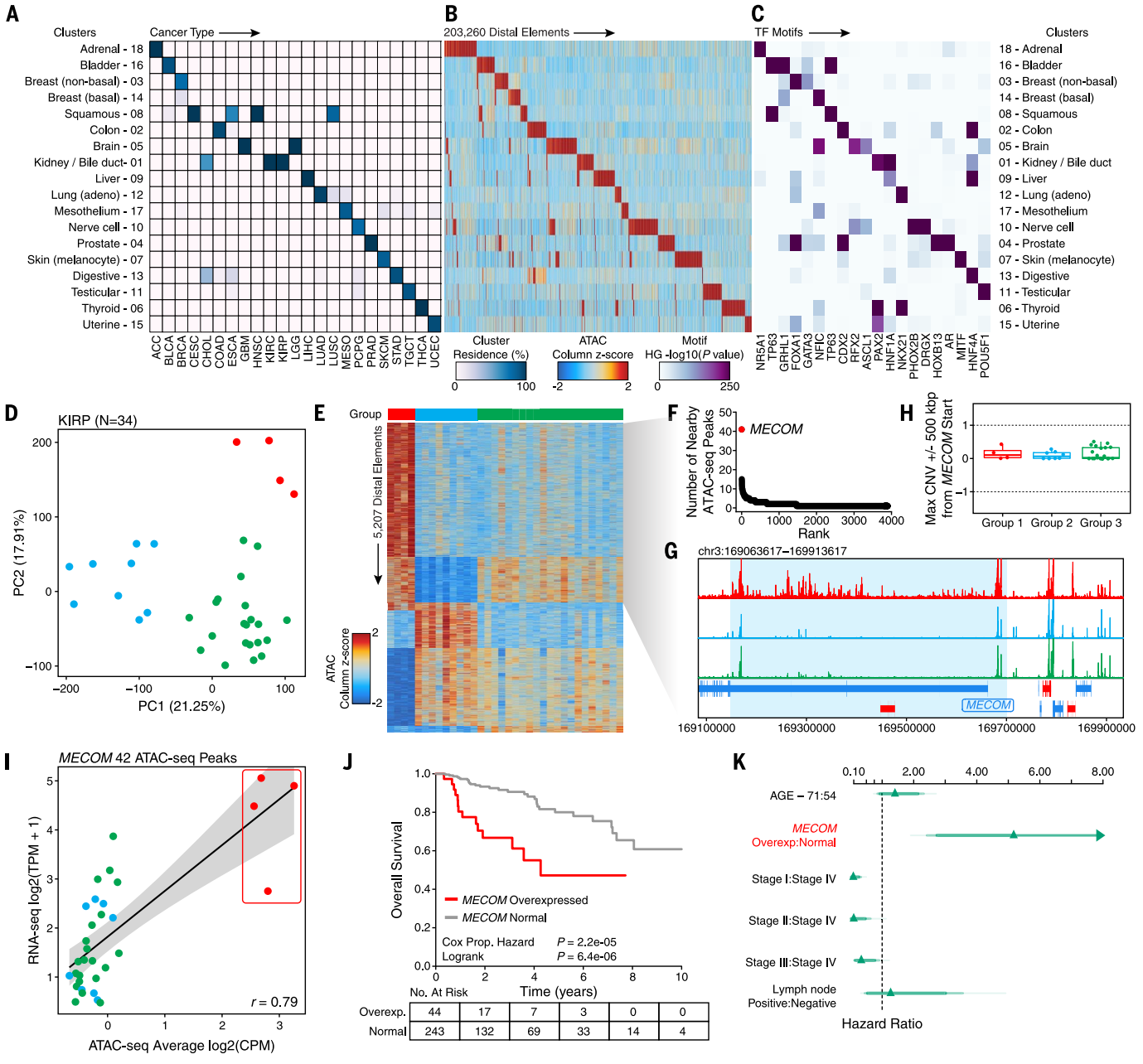


Fig. 3. ATAC-seq clusters cancer samples to show cancer- and tissue-specific drivers. (**A**) Cluster residence heatmap showing the percent of samples from a given cancer type that reside within each of the 18 annotated ATAC-seq clusters. (**B**) Heatmap showing the ATAC-seq accessibility at distal elements (N = 203,260) identified to be cluster-specific by distal binarization. (**C**) Enrichment of TF motifs in peak sets identified in Fig. 3B. Enrichment is determined by a hypergeometric (HG) test $-\log_{10}(P$ value) of the motif's representation within the cluster-specific peaks compared to the pan-cancer peak set. Transcription factors shown represent a manually trimmed set of factors whose expression is highly correlated ($r > 0.4$) with the accessibility of the corresponding motif. Color represents the $-\log_{10}(P$ value) of the hypergeometric test. (**D**) Principal component analysis of the top 25,000 distal ATAC-seq peaks within the KIRP cohort (N = 34 samples). Each dot represents an individual sample. The color of the dots represents K-means clustering (K = 3 by gap statistic). (**E**) Distal binarization analysis based on the three K-means–defined groups identified and shown (by color) in Fig. 3D. (**F**) Dot plot showing the number of nearby ATAC-seq peaks per gene from the group 1 distal binarization. Each dot represents a different gene. The *MECOM* gene (also called *EVI1*) is highlighted in red. (**G**) Normalized average sequencing tracks of K-means–defined groups 1, 2, and 3 at the *MECOM* locus. Peaks specific to group 1 are highlighted by light blue shading. (**H**) DNA copy number data at the *MECOM* locus in the three K-means–defined groups. Each dot represents an individual sample. CNV, copy number variation. (**I**) Average chromatin accessibility at peaks near the *MECOM* gene (N = 42 peaks) and RNA-seq gene expression of *MECOM* in KIRP samples (N = 34 samples). Each dot represents an individual donor. Dots are colored according to the clustering group colors shown in Fig. 3D. CPM, counts per million. (**J**) Kaplan-Meier analysis of overall survival of all KIRP donors in TCGA (N = 287) stratified by *MECOM* overexpressed (N = 44) and normal *MECOM* expression (N = 243). (**K**) Hazard plot of risk of dying from KIRP based on multiple covariates, including *MECOM* expression (hazard ratio = 5.2, 95% confidence interval = 2.4 to 11.0). Lines represent 95% confidence intervals.

accessibility identified by distal binarization (red coloring in Fig. 3E). Within the set of regulatory elements that are specific to this subgroup, we found 42 ATAC-seq peaks near the MDS1 and EVI1 complex locus (*MECOM*) gene (Fig. 3, F and G). Notably, the high chromatin accessibility of these *MECOM* peaks is not related to copy number amplification, as determined by DNA copy number array data (Fig. 3H). The expression of the *MECOM* gene is highly correlated with the mean ATAC-seq accessibility at these 42 ATAC-seq peaks [correlation coefficient (r) = 0.79, Fig. 3I]. Additionally, overexpression of *MECOM* is significantly associated with poorer overall survival across all available KIRP data from TCGA ($P = 2.2 \times 10^{-5}$, Cox proportional hazard test, Fig. 3J) with a hazard ratio of 5.2 (95% confidence interval = 2.4 to 11.0). This association is more substantial than lymph node status or patient age and is independent of cancer stage (Fig. 3K), indicating a potential prognostic role for these findings. Importantly, *MECOM* overexpression is not readily explained by any previously identified subgroups of KIRP, including subgroups with a CpG island methylator phenotype or mutations in the gene encoding fumarate hydratase, which have also been shown to confer poor overall survival (*13*). These results suggest that *MECOM* activation in KIRP identifies a previously unappreciated subgroup of patients with adverse outcomes, a finding that was uncovered by notable changes in the chromatin accessibility landscape of these samples.

Similarly, we found multiple distinct subgroups of PRAD and BRCA based on *K*-means clustering of the top 25,000 variable distal ATAC-seq peaks (fig. S3, C and D). In PRAD, these include subgroups driven by activity of AR, tumor protein P63 (*TP63*), and forkhead box–family TFs (fig. S3C). From an unsupervised analysis of breast cancer, we identified motifs of known TF drivers of luminal subtype identity, such as GATA binding protein 3 (GATA3) and FOXA1, as being enriched in the peak clusters specific to a subset of luminal samples (clusters 3 and 4, fig. S3D). We also identified a potential role for grainyhead-like (GRHL) TF motifs in basal breast cancer (*32*) (cluster 1, fig. S3D) and an overlapping role for nuclear factor I (NFI) in both basal and luminal A breast cancer (cluster 2, fig. S3D). Additionally, ATAC-seq data can be used to identify regions of copy number amplification de novo (*33*), enabling the classification of *HER2*-amplified cases of breast cancer (fig. S3, E to G).

Footprinting analysis defines TF activities in cancer

The high sequencing depth of the ATAC-seq data generated in this study (median of 56.7 million unique reads per technical replicate) enabled the profiling of TF occupancy at base-pair resolution through TF footprinting. TF binding to DNA protects the protein-DNA binding site from transposition while the displacement or depletion of one or more nucleosomes creates high DNA accessibility in the immediate flanking sequence. Collectively, these phenomena are referred to as the TF footprint. To characterize TF footprints, we adapted a recent approach (*34*) that quantifies the "flanking accessibility," a measure of the accessibility of the DNA adjacent to a TF motif, and "footprint depth," a measure of the relative protection of the motif site from transposition (Fig. 4A and data S6). To calculate these variables, we aggregated all insertions relative to the TF motif center, genome-wide (fig. S4A). To attempt to account for known Tn5 transposase insertion bias, we computed the hexamer frequency centered at Tn5 insertions and normalized for the expected bias at each position relative to the motif center (*34*) (see methods for potential limitations). Depending on the binding properties of a TF and its ability to affect local chromatin accessibility, changes in these properties would be detectable through this approach genome-wide (fig. S4, B and C). ChromVAR (*35*), a similar genome-wide approach which assesses the ability of a TF to affect flanking accessibility, identified a highly overlapping list of TFs (fig. S4D).

To uncover transcriptionally driven TF binding patterns, we correlated the RNA-seq gene expression of a given TF to its corresponding footprint depth and estimated flanking accessibility (data S6). A factor whose expression is sufficient to generate robust DNA binding would have a footprint depth and flanking accessibility that are significantly correlated to its gene expression [false discovery rate (FDR) < 0.1, purple dots in Fig. 4B], such as TP63 (Fig. 4, C and D) or NK2 homeobox 1 (NKX2-1) (Fig. 4, E and F). Increases in flanking accessibility and decreases in footprint depth are likewise accompanied by decreases in methylation (bottom of Fig. 4, D and F), consistent with the hypothesis that methylated DNA is less likely to be bound by TFs (*36*). Although footprint depth and flanking accessibility are often correlated, their divergence can suggest the modes of TF-DNA interaction. For example, factors whose expression is sufficient to cause opening of chromatin around the motif site but not to protect the motif site from transposition would be expected to only exhibit a significant correlation between gene expression and flanking accessibility (blue dots in Fig. 4B). This pattern of correlation could be caused by effects such as rapid TF off rates or low occupancy (fig. S4, E and F). Conversely, a small number of TFs have expression that is only significantly correlated with footprint depth (red dots in Fig. 4B). Though likewise rare, we also identified potential negative regulators whose expression is inversely correlated to gain of flanking accessibility and loss of footprint depth, such as the cut-like homeobox 1 (CUX1) TF (*37*) (Fig. 4B and fig. S4, G and H). This is the expected behavior of repressive TFs that bind DNA and lead to compaction of the neighboring sequence. These results predicted dozens of positive and negative regulators whose expression is strongly correlated with chromatin accessibility patterns near to their corresponding motif (fig. S4I and data S6). Overall, our footprinting analysis identified putative TFs with activities correlated with gene expression.

Linking of DNA regulatory elements to genes predicts interactions relevant to cancer biology

The breadth and depth of this sequencing study enabled a robust association of ATAC-seq peaks with the genes that they are predicted to regulate. To do this, we implemented a strategy based on the correlation of ATAC-seq accessibility and gene expression across all samples (Fig. 5A, $N = 373$ with matched RNA-seq and ATAC-seq). Because promoter capture Hi-C data suggested that >75% of three-dimensional (3D) promoter-based interactions occur within a 500–kilobase pair (kbp) distance (*38*), we restricted the length scale of this analysis to 500 kbp to avoid spurious predictions. Using a conservative FDR cutoff of 0.01, we identified 81,323 unique links between distal ATAC-seq peaks and genes (Fig. 5B and data S7). Some of these links are driven by correlation across many cancer types (Fig. 5, C to E), whereas 70% are strongly driven by one cluster (Fig. 5F and data S7). To derive a final list of peak-to-gene links (Fig. 5B), putative links were filtered against (i) links whose correlation is strongly driven by DNA copy number amplification ("CNA"; fig. S5, A and B), (ii) regions with broad and high local correlation ("diffuse"; fig. S5, B and C), and (iii) links involving an ATAC-seq peak that overlaps the promoter of any gene (Fig. 5G). As expected, the histogram of distances between a peak and its target gene decays sharply with distance (*39*) (Fig. 5H). The expression of most genes is correlated with the activity of fewer than five different peaks (Fig. 5I), whereas most peaks are predicted to interact with a single gene (Fig. 5J). Additionally, this analysis found that only 24% of predicted links occur between an ATAC-seq peak and the nearest gene, indicating that the majority of predicted interactions skip over one or more genes and would not be possible to predict from primary sequence alone (Fig. 5K). In total, we predicted at least one peak-to-gene link for 8552 protein-coding genes, accounting for nearly half of all protein-coding genes in the human genome, including 48% of the curated Catalogue of Somatic Mutations in Cancer (COSMIC) cancer-relevant genes (data S7).

In addition to predicting peak-to-gene links across cancer types, we also predicted peak-to-gene links within breast cancer ($N = 74$ donors), identifying 9711 unique peak-to-gene links (fig. S5D and data S7). Of these links, 36% were also identified in our analysis of all cancer types (fig. S5E). Particularly important in these BRCA-specific links was the contribution of recurrent DNA CNA as a strong driver for spurious peak-to-gene correlation (Fig. 5G). These false-positive associations were removed through the use of published TCGA DNA copy number array data and a local correlation correction model, as mentioned above (see methods). The final predicted BRCA-specific links follow a similar distance

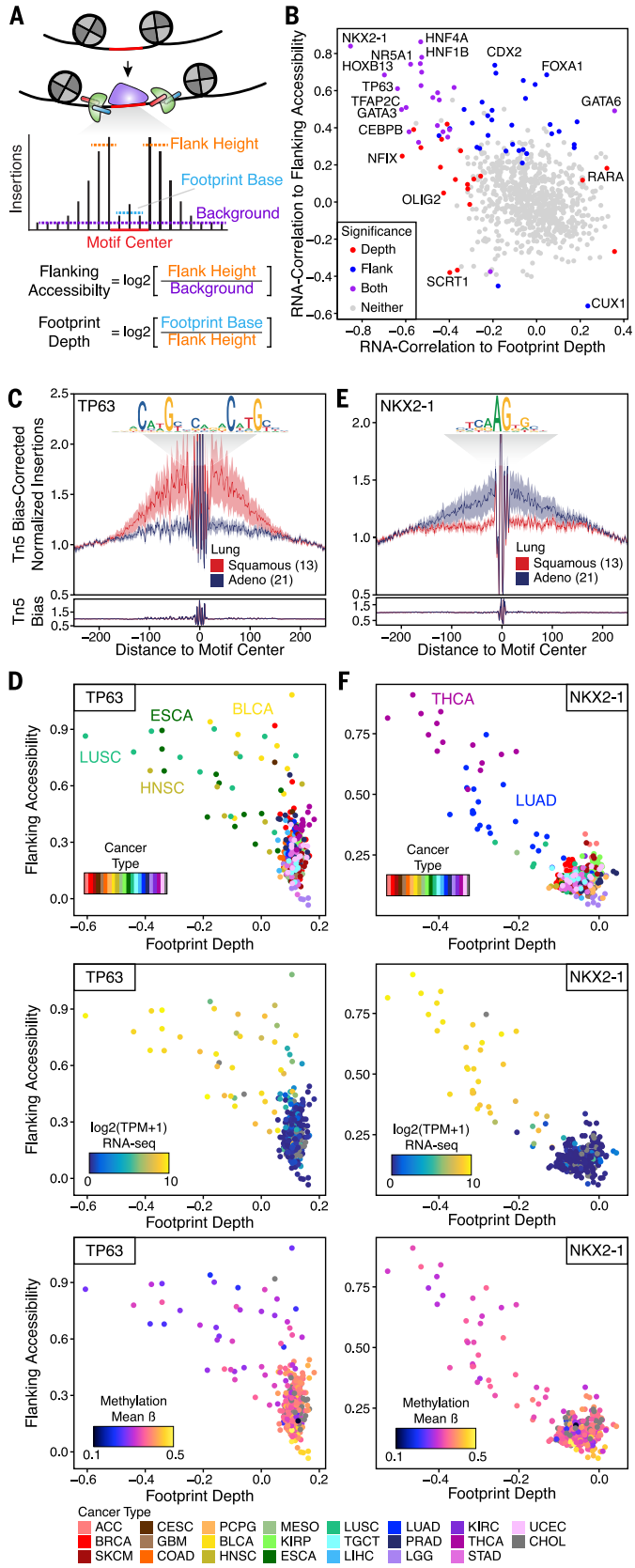


Fig. 4. Footprinting analysis identifies distinct TF activities in cancer. (**A**) Schematic illustrating the dynamics of TF binding (purple) and Tn5 insertion (green). (**B**) Classification of TFs by the correlation of their RNA expression to the footprint depth and flanking accessibility of their motifs. Color represents whether the depth (red), flank (blue), or both (purple) are significantly correlated to TF expression below an FDR cutoff of 0.1. Each dot represents an individual deduplicated TF motif (see methods). (**C**) TF footprinting of the TP63 motif (CIS-BP M2321_1.02) in lung cancer samples from the squamous (cluster 8) or adenocarcinoma (cluster 12) subtype. The Tn5 insertion bias track of TP63 motifs is shown below. (**D**) Dot plots showing the footprint depth and flanking accessibility of TP63 motifs across all lung cancer samples studied. Each dot represents a unique sample. Color represents cancer type (top), RNA-seq gene expression (middle), or methylation beta value (bottom). Samples without matching RNA or methylation data are shown in gray. (**E**) TF footprinting of the NKX2-1 motif (CIS-BP M6374_1.02) in lung cancer samples from the squamous (cluster 8) and adenocarcinoma (cluster 12) subtype. The Tn5 insertion bias track of NKX2-1 motifs is shown below. (**F**) Dot plots showing the footprint depth and flanking accessibility of NKX2-1 motifs across all lung cancer samples studied. Each dot represents a unique sample. Color represents cancer type (top), RNA-seq gene expression (middle), or methylation beta value (bottom). Samples without matching RNA or methylation data are shown in gray.

distribution and peak-to-gene linking specificity as observed in the pan-cancer predicted links (fig. S5, F to I).

Many of these predicted peak-to-gene links occur in clusters where multiple nearby peaks are predicted to be linked to the same gene, indicating that these clusters of peak-to-gene links may function as part of a single regulatory unit or enhancer. Extending the width of the linked ATAC-seq peaks to 1500 bp allows for joining of these peaks into defined merged putative enhancer units (fig. S5J). This resulted in a total of 58,092 pan-cancer and 7622 BRCA-specific enhancer-to-gene links (data S7).

Validation and utility of predicted links between distal elements and genes

To verify a regulatory interaction for the predicted peak-to-gene links, we used a CRISPR interference (CRISPRi) (*40*) strategy using a catalytically dead Cas9 (dCas9) fused to a Kruppel-associated box (KRAB) domain, which mediates focal heterochromatin formation and functional silencing of noncoding DNA regulatory elements (Fig. 6A). In this way, targeting the distal peak region of a predicted peak-to-gene link would be expected to cause a decrease in the expression of the linked gene, located tens to hundreds of kilobases away. CRISPRi of a predicted distal regulatory element linked to *BCL2* (164 kbp, Fig. 5C) led to a significant reduction in *BCL2* gene expression in the luminal-like breast cancer MCF7 cell line but not in the basal-like MDA-MB-231 cell line (Fig. 6B), consistent with the role of *BCL2* as a luminal-specific survival factor (*41*). Similarly, CRISPRi of a distal regulatory element linked to the *SRC* oncogene (−49 kbp, Fig. 5D) led to a significant reduction in gene expression in both MCF7 cells and MDA-MB-231 cells (Fig. 6B). On a genome-wide scale, the predicted BRCA-specific peak-to-gene links show a strong enrichment in 3D chromosome conformation data from MDA-MB-231 cells (*42*), providing further support for our link prediction strategy (Fig. 6C). Moreover, we found that, of the peak-to-gene links predicted from BRCA ATAC-seq data that are also associated with a DNA methylation array CpG probe, 35% overlap with links predicted jointly from DNA methylation array and RNA-seq data in an ELMER analysis (*8*, *43*) of the complete TCGA BRCA dataset (N = 858 tumors) (P << 0.001; Fig. 6D, fig. S6A, and data S8). These overlaps contain many luminal-specific and basal-specific links (fig. S6A), with a clear delineation between luminal (fig. S6B) and basal (fig. S6C) breast cancer samples. Integrating WGBS and ATAC-seq demonstrated the dynamics of methylation and chromatin accessibility and the overlap of predicted interactions at the non-basal *FOXA1* and basal forkhead box C1 (*FOXC1*) loci (fig. S6, D and E).

Similarly, previous work has leveraged TCGA RNA-seq data to infer transcriptional networks that consist of regulons, each of which is based on a TF regulator and its associated positive and negative target genes (fig. S7A) (*44*). For each

regulon, every donor in the cohort can be assigned a positive, undefined, or negative regulon activity as measured by a differential enrichment score (dES) (*45*). Certain patterns of chromatin accessibility are expected on the basis of the target gene set and dES status of the donor (fig. S7B). For example, in donors with positive dES, chromatin at sites linked to positive target genes should be more accessible, whereas chromatin at sites linked to negative targets should be less accessible (fig. S7B). Examination of the estrogen receptor 1 (ESR1) regulon in the 74 BRCA donors profiled in this study identified 482 ATAC-seq distal peak-to-gene links corresponding to 124 ESR1 target genes (fig. S7C and data S8). Accessibility at these peaks is strongly concordant with expectations, further supporting the predicted links ($P < 1 \times 10^{-20}$, fig. S7D). Examination of this regulon across all TCGA BRCA donors (N = 1082) showed a significant difference in overall survival between ESR1 dES-positive and -negative samples (fig. S7, E and F).

Together, pan-cancer and BRCA-specific peak-to-gene links further informed cancer-related GWAS polymorphisms, allowing the linkage of SNPs to putative gene targets with about 65% of all GWAS polymorphisms targeting a gene other than the closest gene on the linear genome (data S5). SNPs falling within peak-to-gene links were predicted to act on important cancer-related genes, including master regulators of cancer and tissue identity such as *NKX2-1* (fig. S7G) and *TP63* (fig. S7H). Focusing specifically on the BRCA peak-to-gene links for which published 3D chromosome conformation data are available, we found clear examples of GWAS SNPs interacting with distant, non-neighboring genes, such as *OSR1* (Fig. 6E and fig. S7I). Moreover, overlapping of the pan-cancer and breast cancer–specific peak-to-gene links with expression quantitative trait loci (eQTLs, where genetic variation at noncoding elements is associated with gene expression differences) from the Genotype-Tissue Expression (GTEx) project showed significant overlap in almost all comparisons (N = 44 of 48 comparisons) (fig. S7J and data S5). These results underscored our ability to use these predicted peak-to-gene links to generate key insights into published data and inform poorly understood aspects of cancer biology.

Identification of DNA regulatory elements related to immunological response to cancer

Of particular interest to current cancer therapy, immune infiltrates represent a substantial

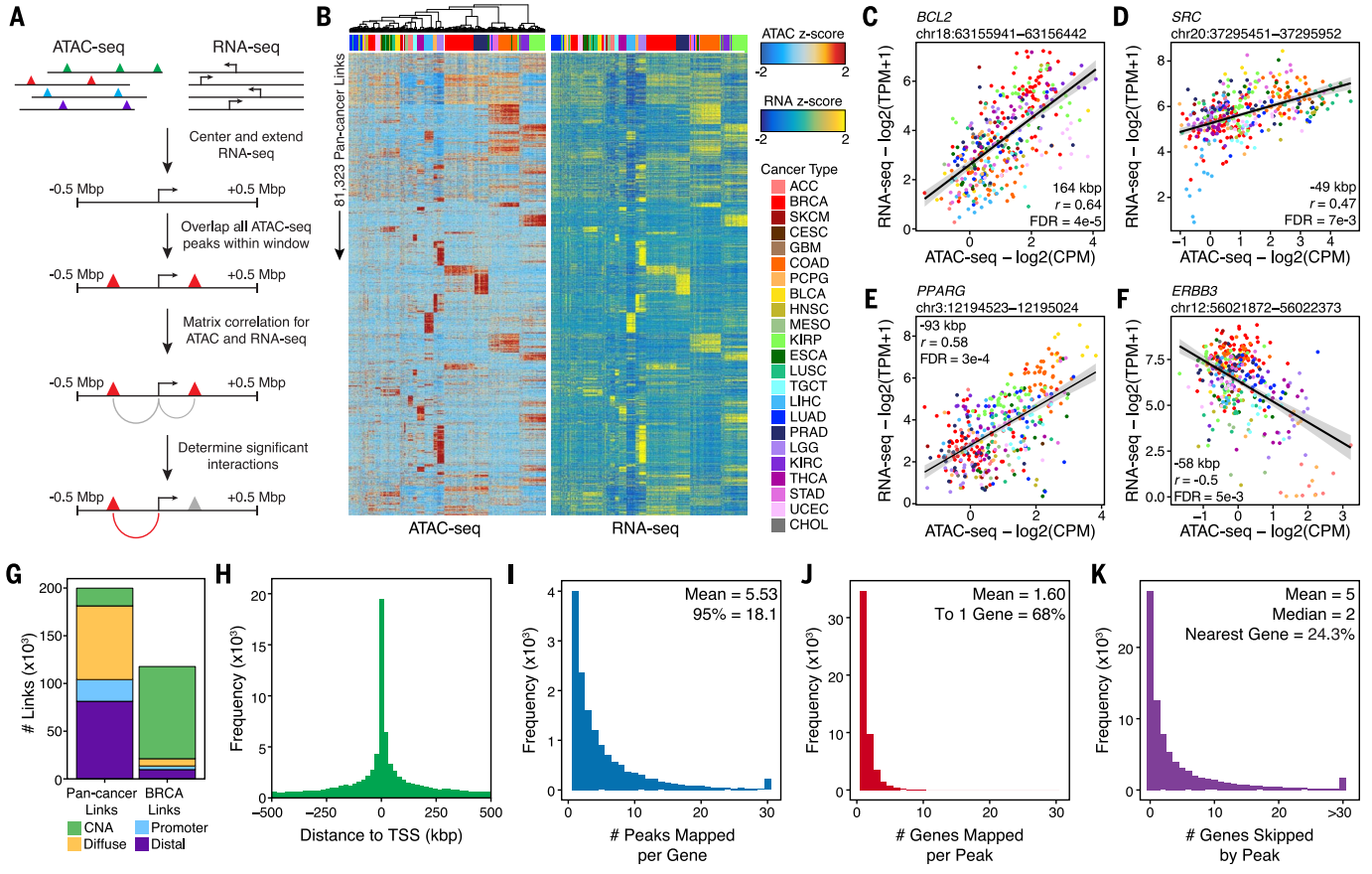


Fig. 5. In silico linking of ATAC-seq peaks to genes. (**A**) Schematic of the in silico approach used to link ATAC-seq peaks in distal noncoding DNA elements to genes via correlation of chromatin accessibility and RNA expression. (**B**) Heatmap representation of the 81,323 pan-cancer peak-to-gene links predicted. Each row represents an individual link between one ATAC-seq peak and one gene. Color represents the relative ATAC-seq accessibility (left) or RNA-seq gene expression (right) for each link as a *z*-score. (**C**) Dot plot of the ATAC-seq accessibility and RNA-seq gene expression of a peak-to-gene link located 164 kbp away from the transcription start site of the *BCL2* gene (peak 498895) that is predicted to regulate its expression. Color represents the cancer type. Each dot represents an individual sample. (**D**) Same as in Fig. 5C but for a peak that is located 49 kbp away from the *SRC* gene (peak 525295). (**E**) Same as in Fig. 5C but for a peak that is located 93 kbp away from the *PPARG* gene (peak 98874). (**F**) Same as in Fig. 5C but for a peak that is located 58 kbp away from the *ERBB3* gene (peak 381116). (**G**) Bar plot showing the number of predicted links that were filtered for various reasons. First, regions whose correlation is driven by DNA copy number amplification were excluded ("CNA"). Next, regions of high local correlation were filtered out ("Diffuse"). Lastly, peak-to-gene links where the peak overlapped a promoter region were excluded ("Promoter"). The remaining links ("Distal") are used in downstream analyses. (**H**) Distribution of the distance of each peak to the transcription start site (TSS) of the linked gene. (**I**) Distribution of the number of peaks linked per gene. (**J**) Distribution of the number of genes linked per peak. (**K**) Distribution of the number of genes "skipped" by a peak to reach its predicted linked gene.

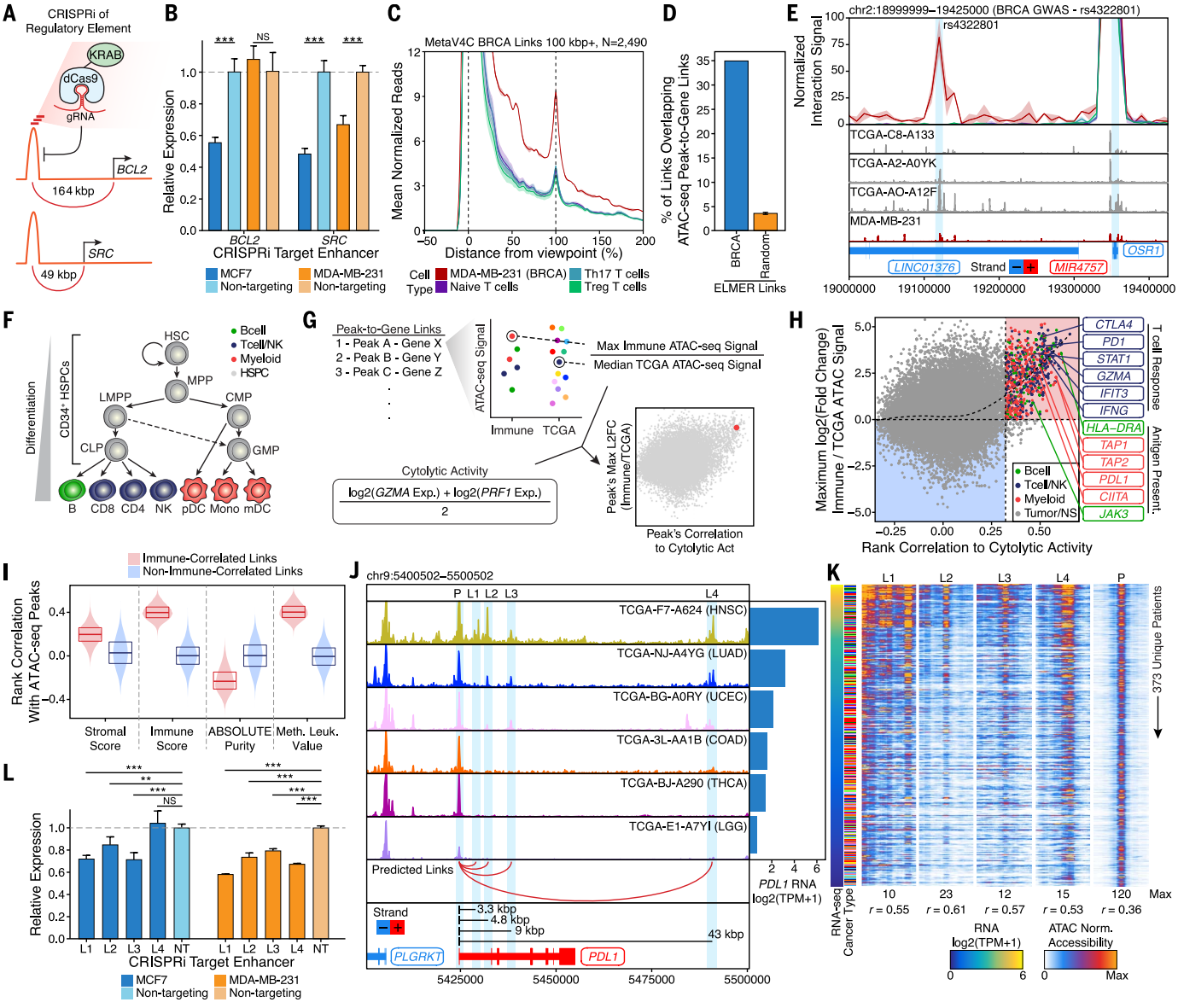
contribution to the overall tumor composition in solid tumors (*46*–*48*). We reasoned that infiltrating immune cells could contribute to our ATAC-seq data, both through actions on tumor cells and through increased chromatin accessibility at known immune-specific regulatory elements. Leveraging published ATAC-seq datasets from the human hematopoietic system (*25*) and data generated here from human dendritic cell subsets (Fig. 6F), we characterized each of our linked peaks by comparing its accessibility in immune cell types to its accessibility in bulk cancer samples (Fig. 6G). We reasoned that peaks that are more accessible in immune cells compared with our cancer cohort might be generated from immune cells associated with the tumor tissue (Fig. 6G). Additionally, we correlated each linked peak to the cytolytic activity score (*49*) of the tumor. The cytolytic activity score is based on the log-average gene expression of granzyme A and perforin 1, two CD8 T cell–specific markers. Linked peaks that exhibit high correlation to cytolytic activity might also be considered to be related to immune infiltration. Combining these two metrics, we identified peak-to-gene links expected to be highly relevant to immune infiltration, including links to genes relevant to antigen presentation and T cell response (Fig. 6H and data S9). The accessibility of these peak-to-gene links that were predicted to be immune-related is highly correlated with computationally predicted metrics of immune infiltration (*46*, *47*) and inversely correlated with tumor purity (*48*) (Fig. 6I). One notable linked gene is programmed death ligand 1 (*PDL1*, also known as *CD274*), a key mediator of immune evasion by cancer and an important target for cancer immunotherapy. *PDL1* is linked to four putative distal regulatory elements that exhibit distinct chromatin accessibility across cancer types and are located as far as 43 kbp away from the *PDL1* transcription start site (Fig. 6, J and K). CRISPRi of each of these four putative *PDL1* regulatory elements significantly decreased, but did not abrogate, the expression of *PDL1* mRNA in at least one of the two breast cancer cell lines tested (MCF7 and MDA-MB-231 cells, Fig. 6L). These results support a model where the expression of *PDL1* is affected by the combined activity of multiple distal regulatory elements.

Identification of cancer-relevant noncoding mutations

In addition to identifying gene regulatory interactions in cancer, ATAC-seq combined with whole-genome sequencing (WGS) can be used to identify regulatory mutations driving cancer initiation and progression. For example, if a noncoding somatic mutation causes the generation of a TF binding site, this mutation could lead to an increase in chromatin accessibility in cis and a concomitant increase in the observed frequency of the mutant allele in ATAC-seq as compared with that in WGS (Fig. 7A). Similarly, a mutation that inactivates a TF binding site can lead to a decrease in chromatin accessibility and a concomitant decrease in the observed frequency of the mutant allele. If such mutations in regulatory elements were to be functional in cancer, we might also expect that they increase or decrease chromatin accessibility beyond the expected distribution observed in nonmutated samples.

From the 404 donors profiled in this study, high-depth WGS data was available for 35 donors across 10 cancer types. These 35 donors had 374,705 called somatic mutations, with 32,696 falling within annotated ATAC-seq peaks and 2259 having at least 30 reads in both ATAC-seq and WGS data (data S10). Among these mutations were three separate occurrences of telomerase reverse transcriptase (*TERT*) gene promoter mutations (Fig. 7B), previously shown to generate de novo E26 transformation-specific (ETS) motif sites. ATAC-seq is especially well suited to identifying these *TERT* promoter mutations because the variant allele frequency is skewed owing to the increase in accessibility on the mutant allele (fig. S8A). Compared with the publicly available exome sequencing data from TCGA, where the *TERT* capture probes do not extend into the promoter region, ATAC-seq provided significantly

Fig. 6. Validation of long-range gene regulation of cancer in peak-to-gene links. (**A**) Schematic of CRISPRi experiments performed. Each experiment uses three guide RNAs (gRNAs) to target an individual peak. The effect of this perturbation on the expression of the linked gene is determined with quantitative polymerase chain reaction (qPCR). (**B**) Gene expression changes by qPCR after CRISPRi of peaks predicted to be linked to the *BCL2* (peak 498895) and *SRC* (peak 525295) genes in MCF7 and MDA-MB-231 cells. Error bars represent the standard deviation of four technical replicates. ***$P < 0.001$ and NS is not significant by two-tailed Student's *t* test. (**C**) Meta-virtual circular chromosome conformation capture (4C) plot of predicted BRCA-specific peak-to-gene links with distances greater than 100 kbp. HiChIP interaction frequency is shown for the MDA-MB-231 basal breast cancer cell line as well as multiple populations of primary T cells. Th17, T helper 17 cell; Treg, regulatory T cell. (**D**) Bar plot showing the overlap of predicted ATAC-seq–based peak-to-gene links and DNA methylation–based ELMER predicted probe-to-gene links in BRCA, as a percentage of all ATAC-seq–based peak-to-gene links with a peak overlapping a methylation probe. The percentage of peak-to-gene links overlapping an ELMER probe-to-gene link (34.9%) is compared to the overlap with 1000 sets of randomized ELMER probe-to-gene links (3.6 ± 0.6%, $P << 0.001$). (**E**) Virtual 4C plot of the peak-to-gene link between rs4322801 and the *OSR1* gene. Normalized HiChIP interaction signal is shown for the MDA-MB-231 basal breast cancer cell line as well as multiple populations of primary T cells using the colors shown in Fig. 6C. ATAC-seq sequencing tracks are shown below for four BRCA samples and MDA-MB-231 cells with increasing levels of *OSR1* gene expression. The rs4322801 SNP (left) and *OSR1* gene (right) are highlighted by light blue shading. Region shown represents chr2:18999999 to 19425000. (**F**) Diagram of the hematopoietic differentiation hierarchy with differentiated cells colored as either B cells (green), T cell or natural killer (Tcell/NK) cells (blue), or myeloid cells (red). HSC, hematopoietic stem cell; LMPP, lymphoid-primed multipotent progenitor; CLP, common lymphoid progenitor; MPP, multipotent progenitor; CMP, common myeloid progenitor, GMP, granulocyte macrophage progenitor; HSPC, hematopoietic stem and progenitor cells; pDC, plasmacytoid dendritic cell; mDC, myeloid dendritic cell. (**G**) Schematic of the analysis shown in Fig. 6H. Peak-to-gene links are classified as related to immune infiltration if their accessibility is higher in immune cells than TCGA cancer samples and they are highly correlated to cytolytic activity. (**H**) Dot plot showing ATAC-seq peak-to-gene links with relevance to immune infiltration. Each dot represents an individual peak with a predicted gene link. Peaks that are related to immune cells have higher ATAC-seq accessibility in immune cell types compared to TCGA cancer samples. Peaks related to immune infiltration have a higher correlation to cytolytic activity. Color represents the cell type of the observation. The vertical dotted line represents the mean + 2.5 standard deviations above the mean for all ATAC-seq peak correlations to the cytolytic activity. The red shading indicates peak-to-gene links that are predicted to be related to immune infiltration. The blue shading indicates peak-to-gene links that are not predicted to be related to immune infiltration. NS, not significant. (**I**) Violin plots of the distribution of Spearman correlations across all peak-to-gene links predicted to be related to immune infiltration (red) or not (blue) with various metrics of tumor purity. (**J**) Normalized ATAC-seq sequencing tracks of the *PDL1* gene locus in six samples with variable levels of expression of the *PDL1* gene (right). Predicted links (red) are shown below for four peak-to-gene links (L1 to L4, peaks 293734, 293735, 293736, and 293740, respectively) to the promoter of *PDL1*. One of these peak-to-gene links (L2) overlaps an alternative start site for *PDL1* and was therefore labeled as a "promoter" peak during filtration. This peak-to-gene link was added to this analysis after manual observation. Region shown represents chr9:5400502 to 5500502. (**K**) Heatmap representation of the ATAC-seq chromatin accessibility of the 5000-bp region centered at each of the four peak-to-gene links shown in Fig. 6J. Each row represents a unique donor ($N = 373$) ranked by *PDL1* expression. The correlation of the chromatin accessibility of each peak with the expression of *PDL1* is shown below the plot. Color represents normalized accessibility. (**L**) Gene expression changes of the *PDL1* gene by qPCR after CRISPRi of peaks predicted to be linked to the *PDL1* gene in MCF7 and MDA-MB-231 cells. Error bars represent the standard deviation of four technical replicates. ***$P < 0.0001$ and **$P < 0.05$ by two-tailed Student's *t* test.



higher sequencing coverage of the *TERT* promoter locus per read sequenced, enabling a more robust classification of *TERT* promoter mutations ($P < 1 \times 10^{-7}$, fig. S8B). Of the three *TERT* promoter mutations identified in the subset of donors with matched WGS, one mutation, in particular, leads to a significant increase in accessibility compared to the other nonmutated members of that cancer type (FDR < 0.0001, blue dot in Fig. 7, B and C). As expected, this increase in *TERT* promoter accessibility is associated with a concomitant increase in *TERT* gene expression (blue dot in Fig. 7C). *TERT* promoter mutations, however, are not the only way to increase *TERT* gene expression, because high *TERT* expression can also be observed in samples without identifiable *TERT* promoter mutations (Fig. 7C). Consistent with a previous report (*50*), differential motif analysis at the site of this *TERT* promoter mutation identified E74-like ETS transcription factor 1 (ELF1) or ELF2 as the TF that likely binds to the de novo ETS motif (fig. S8C). In addition, we identified several mutations overlapping CCCTC-binding factor (CTCF) motif occurrences that are associated with decreased accessibility at that site (fig. S8, D and E). However, these mutations were relatively rare and often had only small effects on the accessibility of the CTCF motif site despite a known enrichment of somatic mutations in CTCF motif sites in cancer (*51*, *52*).

In addition to known *TERT* promoter mutations, integrative analysis of WGS and ATAC-seq data uncovered a mutation upstream of the FYVE RhoGEF and PH domain-containing 4 (*FGD4*) gene, a regulator of the actin cytoskeleton and cell shape. This mutation occurs in a bladder cancer sample where the variant allele frequency observed in ATAC-seq is markedly higher than the variant allele frequency observed in WGS (Fig. 7B). This mutation is associated with a significant increase in accessibility compared to other bladder cancer samples in this cohort (Fig. 7, B and D) and is accompanied by a similar increase in *FGD4* mRNA (Fig. 7D). Moreover, this mutation upstream of the *FGD4* gene (referred to as eFGD4 for enhancer FGD4) leads to a level of accessibility that is higher than any of the other samples profiled by ATAC-seq in this study (fig. S8F) and a level of *FGD4* gene expression that is in the top 3% of all bladder cancer samples in TCGA (fig. S8G). As estimated by WGS data, this eFGD4 mutation is present in a subclone comprising about 13% of the tumor (Fig. 7E); however, the mutant allele is present in 96% of all ATAC-seq reads spanning this locus (Fig. 7E), demonstrating a strong preference for accessibility on the mutant allele. This eFGD4 mutation is analogous to, but potentially more potent than, the *TERT* promoter mutation

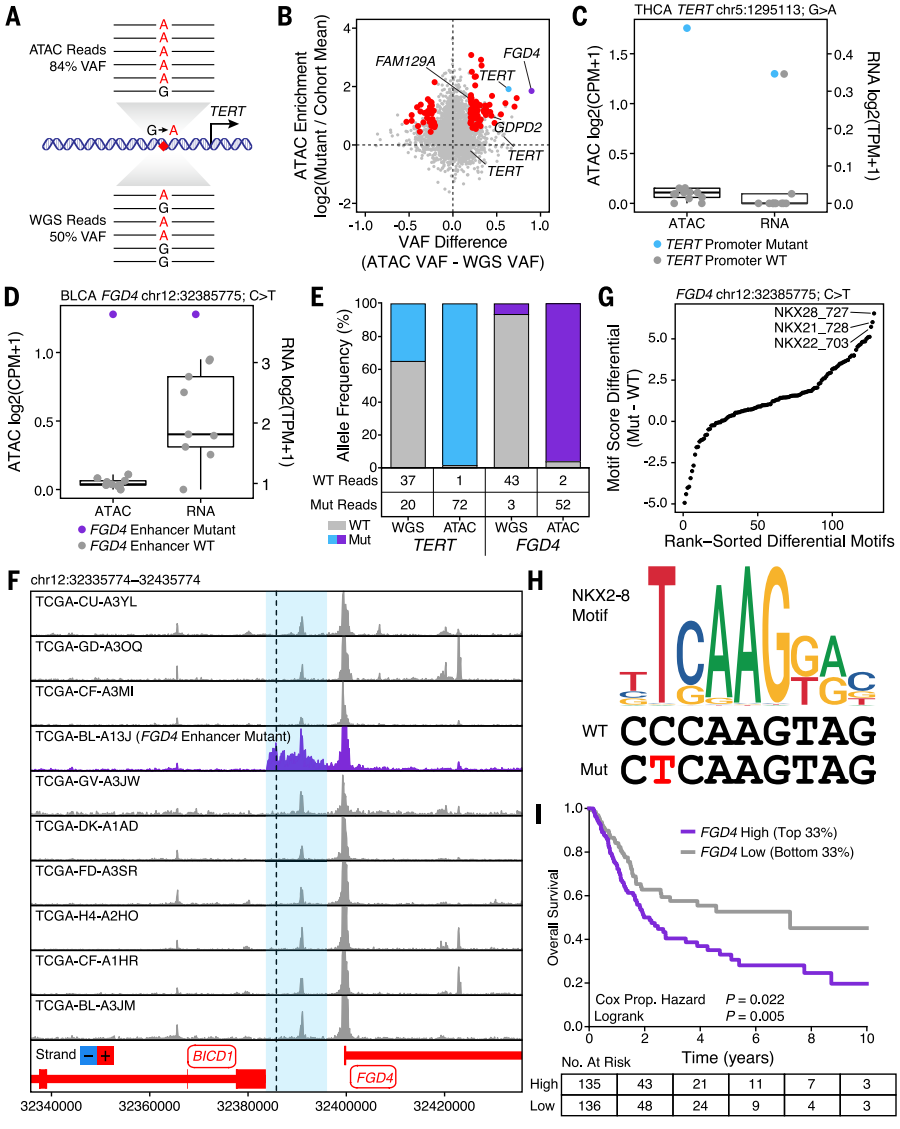


Fig. 7. Integration of WGS and ATAC-seq identifies cancer-relevant regulatory mutations. (**A**) Schematic of how functional variants are identified in regulatory elements. The example shown depicts the *TERT* promoter. (**B**) Dot plot of the difference in variant allele frequency (VAF) of ATAC-seq and WGS and the changes in chromatin accessibility caused by the given variant with respect to other samples of the same cancer type. Variants with a higher variant allele frequency in ATAC-seq than WGS would be expected to cause an increase in accessibility. Each dot represents an individual somatic mutation. (**C**) Normalized ATAC-seq and RNA-seq of thyroid cancer donors profiled in this study. Each dot represents an individual donor. Blue dot represents the donor with a *TERT* promoter mutation shown in Fig. 7B. Other thyroid cancer donors known to harbor a *TERT* promoter mutation were excluded from this plot. The hinges of the box represent the 25th to 75th percentile. WT, wild type. (**D**) Normalized ATAC-seq and RNA-seq of bladder cancer donors profiled in this study. Each dot represents an individual donor. Purple dot represents the donor with a mutation upstream of the *FGD4* gene shown in Fig. 7B. The hinges of the box represent the 25th to 75th percentile. (**E**) Comparison of wild-type and mutant reads in WGS and ATAC-seq data at the *TERT* promoter and *FGD4* upstream region for the donors highlighted in (D) and (E). (**F**) Normalized ATAC-seq sequencing tracks of the *FGD4* locus in the 10 bladder cancer samples profiled in this study, including the one sample with a mutation predicted to generate a de novo NKX motif (TCGA-BL-A13J). Locus shown represents chr12:32335774 to 32435774. The mutation position is indicated by a black dotted line. The predicted enhancer region surrounding this mutation is highlighted by light blue shading. (**G**) Difference in motif score in the wild-type and mutant *FGD4* upstream region. Motif score represents the degree of similarity between the sequence of interest and the relevant motif. Each dot represents an individual motif. (**H**) Overlay of the NXK2-8 motif (CIS-BP M6377_1.02) and the wild-type and mutant sequences of the *FGD4* upstream region. (**I**) Kaplan-Meier survival analysis of TCGA bladder cancer patients with high (top 33%) and low (bottom 33%) expression for the *FGD4* gene.

described above (Fig. 7E). In the case of the eFGD4 mutation, this dramatic allele bias occurs because chromatin at this locus is not normally accessible in any of the bladder cancer samples profiled in this study (gray dots and tracks in Fig. 7, D and F) but becomes highly accessible in the context of the eFGD4 mutation (purple dot and track in Fig. 7, D and F). Differential motif analysis identified NKX factor motifs as the most strongly enriched in the sequence corresponding to the eFGD4 mutation (Fig. 7G), where a C-to-T transition at position two generated a perfect NKX2-8 motif de novo from a latent site (Fig. 7H). RNA-seq data from the mutated sample identified multiple expressed NKX TFs [transcripts per million (TPM) > 0.5], nominating NKX3-1, NKX2-3, and NKX2-5 as potential mediators of this DNA binding event (fig. S8H). From this, we hypothesized that the eFGD4 mutation creates a de novo binding site for an NKX TF which, upon binding to the DNA, leads to a broad increase in accessibility across the entire 12-kbp region upstream of the *FGD4* gene. This hypothesis was further supported by the observation that the ATAC-seq accessibility of the entire *FGD4* upstream locus occurs on a single phased allele (fig. S8I). Moreover, separation of subnucleosomal and nucleosome-spanning reads in the ATAC-seq data are consistent with protein binding at the site of the eFGD4 mutation (light blue shading in fig. S8I). Lastly, because higher *FGD4* expression is significantly associated with worse overall survival in bladder cancer (Fig. 7I and fig. S8J), this mutation could have functional consequence in this particular cancer. Whether the eFGD4 mutation or other enhancer mutations emerge as recurrent drivers of human cancer should be addressed in future studies. Our data identified multiple additional noncoding mutations associated with a concomitant gain of chromatin accessibility and increase in RNA expression (fig. S8, K to Q), and we anticipate that future work will uncover mechanisms underlying this type of regulatory mutation across all cancer types.

Discussion

Here we provide an initial characterization of the chromatin regulatory landscape in primary human cancers. This dataset identified hundreds of thousands of accessible DNA elements, expanding the dictionary of regulatory elements discovered through previous large-scale efforts such as The Roadmap Epigenomics Project. The identification of these additional elements was made possible through (i) our analysis of primary cancer specimens, (ii) greater saturation of some cancer and tissue types in our dataset, or (iii) potential differences between ATAC-seq and DNase-seq platforms. Nevertheless, the high overlap between the two datasets demonstrates the robustness of both platforms and the consistency of the observed results.

The exquisite cell type–specificity of distal regulatory elements from our ATAC-seq data enabled the classification of cancer types and the discovery of previously unappreciated cancer

subtypes. De novo clustering of TCGA samples based on chromatin accessibility strongly overlaps previous integrative clustering methods, identifying 18 distinct cancer clusters. Comparing this clustering scheme to other clustering schemes defined by cancer type, mRNA, miRNA, DNA methylation, RPPA, and DNA copy number alterations, we observed the strongest concordance of our clustering scheme with mRNA and cancer type, consistent with a close functional linkage between chromatin accessibility and transcriptional output. The strength of the observed associations is influenced by the features represented for each platform. For example, the DNA methylation clusters are based on cancer-specific promoter hypermethylation (*28*). Clustering based on DNA methylation at distal regulatory elements would likely show a stronger correlation with the ATAC-seq groupings, but distal regulatory element representation on the DNA methylation array used for these samples was too sparse to allow such an analysis. We also identified epigenetically distinct subtypes of kidney renal papillary cancer that have clear differences in overall survival. This cancer type–specific activity in DNA regulatory elements may arise via mutations within the regulatory element, pathologic transcription factor activity, or reflect the regulatory state of the tumor's cell of origin (e.g., stem cells). As the chromatin accessibility landscapes of additional primary cancer samples are profiled, we anticipate the identification of further epigenetic subdivisions with prognostic implications, potentially nominating avenues for therapeutic intervention.

The data generated in this study fully represents the cellular complexity of primary human tumors, comprising signals from tumor cells, infiltrating immune cells, stromal cells, and other normal cell types. In many ways, this complexity is advantageous because it allows complex systems-level analyses to be performed in the future, including cellular deconvolution approaches to understand the contributions of various cell types or cell states to the overall landscape of chromatin accessibility. However, the admixed nature of this signal also highlights the need for future work to profile the chromatin accessibility of matched healthy tissues to further refine the specific changes that drive cancer. Nevertheless, the chromatin accessibility profiles generated in this study represent the largest effort to date to characterize the regulatory landscape in primary human cancer cells.

Using this data-rich resource, we identified classes of TFs whose expression leads to different patterns in TF occupancy and motif protection. By integrating RNA-seq and ATAC-seq, we found factors whose expression is sufficient for both motif protection and nucleosome repositioning and demonstrated this binding to be inversely correlated with the level of DNA methylation at those binding sites. Despite this strong correlation, many sites of differential chromatin accessibility do not show differential methylation, demonstrating the complementarity of these two data types, perhaps owing to the presence of intermediate chromatin states such as poised promoters or enhancers (*53*, *54*).

Moreover, integration of RNA-seq and ATAC-seq across the 373 donors with paired datasets enabled a quantitative model to link the accessibility of a regulatory element to the expression of predicted target genes. This workflow identified putative links for more than half of the protein-coding genes in the genome, informing the target genes of poorly understood GWAS SNPs and increasing our understanding of cancer gene regulatory networks. These predictions were further supported using 3D chromosome conformation data, and a subset were validated through CRISPRi experiments in breast cancer cell lines. However, profiling of chromosome conformation in primary cancer samples has not been performed on a large scale. Future work to produce maps of chromosome conformation in these or other primary cancer samples will improve our understanding of gene regulatory networks in cancer and further clarify the roles for certain GWAS-identified SNPs in cancer initiation and progression.

Lastly, through integration of WGS and ATAC-seq, we revealed a class of somatic mutations that occur in regulatory regions and lead to strong gains in chromatin accessibility. We demonstrated that these mutations likely lead to changes in nearby gene expression and affect genes whose expression is linked to poorer overall survival. Some of these mutations, such as those occurring in the *TERT* promoter, have been found to be recurrent whereas others, such as the mutation upstream of the *FGD4* gene, may be rare but functionally important. Because the enhancer functions are often distributed and latent enhancer sequences are pervasive in the genome, noncoding mutations in cancer may be especially challenging and require high-throughput functional assessment. Future larger-scale efforts to combine genome and epigenome sequencing will pave the way to tackling the noncoding genome in cancer.

Materials and methods summary

ATAC-seq data was generated from 410 tissue samples from the TCGA collection of primary human tumors. These samples spanned 23 different tumor types. These ATAC-seq data were used to cluster samples, identifying epigenetically defined patient subgroups. Moreover, TF regulators of cancer were defined, and footprinting of these regulators was correlated to gene expression to identify putative classes of TFs. A correlation-based model was developed to link ATAC-seq peaks to putative target genes. These putative links were validated using CRISPRi-based perturbation of the peak region followed by quantification of changes in gene expression. Publicly available HiChIP data and GTEx eQTL data were further used to support genome-wide peak-to-gene linkage predictions. Lastly, WGS and ATAC-seq were combined to identify noncoding mutations that affect chromatin accessibility in an allele-specific manner.

REFERENCES AND NOTES

1. C. Hutter, J. C. Zenklusen, The Cancer Genome Atlas: Creating lasting value beyond its data. *Cell* **173**, 283–285 (2018). doi: 10.1016/j.cell.2018.03.042; pmid: 29625045
2. W. A. Flavahan, E. Gaskell, B. E. Bernstein, Epigenetic plasticity and the hallmarks of cancer. *Science* **357**, eaal2380 (2017). doi: 10.1126/science.aal2380; pmid: 28729483
3. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011). doi: 10.1016/j.cell.2011.02.013; pmid: 21376230
4. M. Egeblad, E. S. Nakasone, Z. Werb, Tumors as organs: Complex tissues that interface with the entire organism. *Dev. Cell* **18**, 884–901 (2010). doi: 10.1016/j.devcel.2010.05.012; pmid: 20627072
5. W. Zhou *et al*., DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat. Genet.* **50**, 591–602 (2018). doi: 10.1038/s41588-018-0073-4; pmid: 29610480
6. M. Almamun *et al*., Integrated methylome and transcriptome analysis reveals novel regulatory elements in pediatric acute lymphoblastic leukemia. *Epigenetics* **10**, 882–890 (2015). doi: 10.1080/15592294.2015.1078050; pmid: 26308964
7. Y. He *et al*., Improved regulatory element prediction based on tissue-specific local epigenomic signatures. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E1633–E1640 (2017). doi: 10.1073/pnas.1618353114; pmid: 28193886
8. L. Yao, H. Shen, P. W. Laird, P. J. Farnham, B. P. Berman, Inferring regulatory element landscapes and transcription factor networks from cancer methylomes. *Genome Biol.* **16**, 105 (2015). doi: 10.1186/s13059-015-0668-3; pmid: 25994056
9. M. Ceccarelli *et al*., Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell* **164**, 550–563 (2016). doi: 10.1016/j.cell.2015.12.028; pmid: 26824661
10. H. Noushmehr *et al*., Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* **17**, 510–522 (2010). doi: 10.1016/j.ccr.2010.03.017; pmid: 20399149
11. T. Hinoue *et al*., Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Res.* **22**, 271–282 (2012). doi: 10.1101/gr.117523.110; pmid: 21659424
12. P. A. Northcott *et al*., The whole-genome landscape of medulloblastoma subtypes. *Nature* **547**, 311–317 (2017). doi: 10.1038/nature22973; pmid: 28726821
13. The Cancer Genome Atlas Research Network, Comprehensive molecular characterization of papillary renal-cell carcinoma. *N. Engl. J. Med.* **374**, 135–145 (2016). doi: 10.1056/NEJMoa1505917; pmid: 26536169
14. B. Akhtar-Zaidi *et al*., Epigenomic enhancer profiling defines a signature of colon cancer. *Science* **336**, 736–739 (2012). doi: 10.1126/science.1217277; pmid: 22499810
15. H. Chen *et al*., A pan-cancer analysis of enhancer expression in nearly 9000 patient samples. *Cell* **173**, 386–399.e12 (2018). doi: 10.1016/j.cell.2018.03.027; pmid: 29625054
16. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013). doi: 10.1038/nmeth.2688; pmid: 24097267
17. M. R. Corces *et al*., An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017). doi: 10.1038/nmeth.4396; pmid: 28846090
18. A. Kundaje *et al*., Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015). doi: 10.1038/nature14248; pmid: 25693563
19. J. Schuijers *et al*., Transcriptional dysregulation of *MYC* reveals common enhancer-docking mechanism. *Cell Reports* **23**, 349–360 (2018). doi: 10.1016/j.celrep.2018.03.056; pmid: 29641996
20. G. Andrey *et al*., A switch between topological domains underlies *HoxD* genes collinearity in mouse limbs. *Science* **340**, 1234167 (2013). doi: 10.1126/science.1234167; pmid: 23744951
21. M. Yeager *et al*., Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.* **39**, 645–649 (2007). doi: 10.1038/ng2022; pmid: 17401363
22. I. P. M. Tomlinson *et al*., A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat. Genet.* **40**, 623–630 (2008). doi: 10.1038/ng.111; pmid: 18372905

23. I. K. Sur *et al*., Mice lacking a *Myc* enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. *Science* **338**, 1360–1363 (2012). doi: 10.1126/science.1228606; pmid: 23118011
24. R. E. Thurman *et al*., The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012). doi: 10.1038/nature11232; pmid: 22955617
25. M. R. Corces *et al*., Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203 (2016). doi: 10.1038/ng.3646; pmid: 27526324
26. L. J. P. van der Maaten, G. E. Hinton, Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
27. A. Rodriguez, A. Laio, Clustering by fast search and find of desity peaks. *Science* **344**, 1492–1496 (2014). doi: 10.1126/science.1242072; pmid: 24970081
28. K. A. Hoadley *et al*., Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell* **173**, 291–304.e6 (2018). doi: 10.1016/j.cell.2018.03.022; pmid: 29625048
29. M. Kulis *et al*., Whole-genome fingerprint of the DNA methylome during human B cell differentiation. *Nat. Genet.* **47**, 746–756 (2015). doi: 10.1038/ng.3291; pmid: 26053498
30. H. S. Kim *et al*., Pluripotency factors functionally premark cell-type-restricted enhancers in ES cells. *Nature* **556**, 510–514 (2018). doi: 10.1038/s41586-018-0048-8; pmid: 29670286
31. H. Shen *et al*., Integrated molecular characterization of testicular germ cell tumors. *Cell Reports* **23**, 3392–3406 (2018). doi: 10.1016/j.celrep.2018.05.039; pmid: 29898407
32. S. Werner *et al*., Dual roles of the transcription factor grainyhead-like 2 (GRHL2) in breast cancer. *J. Biol. Chem.* **288**, 22993–23008 (2013). doi: 10.1074/jbc.M113.456293; pmid: 23814079
33. S. K. Denny *et al*., Nfib promotes metastasis through a widespread increase in chromatin accessibility. *Cell* **166**, 328–342 (2016). doi: 10.1016/j.cell.2016.05.052; pmid: 27374332
34. S. Baek, I. Goldstein, G. L. Hager, Bivariate genomic footprinting detects changes in transcription factor activity. *Cell Reports* **19**, 1710–1722 (2017). doi: 10.1016/j.celrep.2017.05.003; pmid: 28538187
35. A. N. Schep, B. Wu, J. D. Buenrostro, W. J. Greenleaf, ChromVAR: Inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017). doi: 10.1038/nmeth.4401; pmid: 28825706
36. Y. Yin *et al*., Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, eaaj2239 (2017). doi: 10.1126/science.aaj2239; pmid: 28473536
37. T. Ellis *et al*., The transcriptional repressor CDP (Cutl1) is essential for epithelial cell differentiation of the lung and the hair follicle. *Genes Dev.* **15**, 2307–2319 (2001). doi: 10.1101/gad.200101; pmid: 11544187
38. B. M. Javierre *et al*., Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* **167**, 1369–1384.e19 (2016). doi: 10.1016/j.cell.2016.09.037; pmid: 27863249
39. C. P. Fulco *et al*., Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* **354**, 769–773 (2016). doi: 10.1126/science.aag2445; pmid: 27708057
40. L. S. Qi *et al*., Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013). doi: 10.1016/j.cell.2013.02.022; pmid: 23452860
41. Y. H. Eom, H. S. Kim, A. Lee, B. J. Song, B. J. Chae, BCL2 as a subtype-specific prognostic marker for breast cancer. *J. Breast Cancer* **19**, 252–260 (2016). doi: 10.4048/jbc.2016.19.3.252; pmid: 27721874
42. S. W. Cho *et al*., Promoter of lncRNA gene *PVT1* is a tumor-suppressor DNA boundary element. *Cell* **173**, 1398–1412.e22 (2018). doi: 10.1016/j.cell.2018.03.068; pmid: 29731168
43. T. C. Silva, S. G. Coetzee, L. Yao, D. J. Hazelett, H. Noushmehr, B. P. Berman, Enhancer linking by methylation/expression relationships with the R package ELMER version 2. bioRxiv 148726 [Preprint]. 11 June 2017. doi: 10.1101/148726
44. A. G. Robertson *et al*., Comprehensive molecular characterization of muscle-invasive bladder cancer. *Cell* **171**, 540–556.e25 (2017). doi: 10.1016/j.cell.2017.09.007; pmid: 28988769
45. M. A. A. Castro *et al*., Regulators of genetic risk of breast cancer identified by integrative network analysis. *Nat. Genet.* **48**, 12–21 (2016). doi: 10.1038/ng.3458; pmid: 26618344
46. V. Thorsson *et al*., The immune landscape of cancer. *Immunity* **48**, 812–830.e14 (2018). doi: 10.1016/j.immuni.2018.03.023; pmid: 29628290
47. K. Yoshihara *et al*., Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013). doi: 10.1038/ncomms3612; pmid: 24113773
48. S. L. Carter *et al*., Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012). doi: 10.1038/nbt.2203; pmid: 22544022
49. M. S. Rooney, S. A. Shukla, C. J. Wu, G. Getz, N. Hacohen, Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61 (2015). doi: 10.1016/j.cell.2014.12.033; pmid: 25594174
50. M. M. Makowski *et al*., An interaction proteomics survey of transcription factor binding at recurrent TERT promoter mutations. *Proteomics* **16**, 417–426 (2016). doi: 10.1002/pmic.201500327; pmid: 26553150
51. R. Katainen *et al*., CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat. Genet.* **47**, 818–821 (2015). doi: 10.1038/ng.3335; pmid: 26053496
52. D. Hnisz *et al*., Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454–1458 (2016). doi: 10.1126/science.aad9024; pmid: 26940867
53. R. D. Hawkins *et al*., Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* **6**, 479–491 (2010). doi: 10.1016/j.stem.2010.03.018; pmid: 20452322
54. T. K. Kelly *et al*., Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* **22**, 2497–2506 (2012). doi: 10.1101/gr.143008.112; pmid: 22960375

ACKNOWLEDGMENTS

We thank X. Ji and J. Coller for assistance in sequencing, P. Giresi and Epinomics for sharing advice and expertise related to ATAC-seq data analysis, and the members of the Greenleaf and Chang laboratories for thoughtful advice and critique. **Funding:** Supported by the National Cancer Institute, NIH grants R35-CA209919 (to H.Y.C.), P50-HG007735 (to H.Y.C. and W.J.G.), and the Parker Institute for Cancer Immunotherapy (H.Y.C.). M.R.C. is supported by NIH K99-AG059918. Additional support through the NIH Genomic Data Analysis Networks 1U24CA210974-01 (J. Zhu), 1U24CA210949-01 (J. N. Weinstein), 1U24CA210978-01 (R. Beroukhim), 1U24CA210952-01 (S. J. Jones), 1U24CA210989-01 (O. Elemento), 1U24CA210990-01 (J. Stuart), 1U24CA210950-01 (R. Akbani), 1U24CA210969-01 (P.W.L.), and 1U24CA210988-01 (K.A.H.). W.J.G. is a Chan-Zuckerberg Biohub Investigator. H.Y.C. is an Investigator of the Howard Hughes Medical Institute. **Author contributions:** L.M.S., J.C.Z., W.J.G., and H.Y.C. conceived of and designed the project. M.R.C. and J.M.G. compiled figures and wrote the manuscript with the help of all authors. M.R.C. developed methodology for profiling frozen cancer tissues by ATAC-seq. S.S., B.H.L., and M.R.C. performed all tissue processing and ATAC-seq data generation. N.C.S. designed and wrote the ATAC-seq data processing pipeline with help from M.R.C. M.R.C. and J.M.G. processed all ATAC-seq data, and J.M.G. performed all analyses and developed all analytical tools unless otherwise stated below. J.A.S. and C.G. performed survival analyses with supervision from C.C., A.G.R., and M.A.C. J.A.S. performed subtyping analysis for KIRP. W.Z. performed WGBS methylation analysis and variation of information clustering analysis with supervision from P.W.L. T.C.S. performed all ELMER analyses with supervision from B.P.B. C.G. performed all regulon analysis with supervision from A.G.R. and M.A.C. C.K.W. performed tumor map analysis. K.A.H. performed cluster coincidence analysis comparing ATAC-seq–derived clusters to TCGA iClusters. S.W.C. produced all Tn5 transposase used in this study and generated reagents and cell lines used in CRISPRi experiments. B.H.L., S.S., and M.R.C. performed CRISPRi experiments. A.T.S. generated human dendritic cell ATAC-seq data. J.M.G. and A.T.S. performed immune infiltration analysis. J.M.G. and M.R.M. performed HiChIP analysis. M.R.C. performed all analysis for identifying noncoding mutations from WGS and ATAC-seq data. I.F. coordinated all TCGA analysis working group efforts. J.C.Z. selected tumor samples to profile in this study. P.W.L. and W.J.G. co-chaired the TCGA analysis working group. C.C. provided expertise relevant to pan-cancer data analysis. H.Y.C. and W.J.G. supervised overall data generation and analysis. All authors listed under "The Cancer Genome Atlas Analysis Network" provided valuable input and expertise. **Competing interests:** H.Y.C. is a co-founder of Accent Therapeutics and Epinomics and is an adviser of 10X Genomics and Spring Discovery. W.J.G is a co-founder of Epinomics and an adviser to 10X Genomics, Guardant Health, Centrillion, and NuGen. C.C. is an adviser to GRAIL. Stanford University holds a patent on ATAC-seq, on which H.Y.C. and W.J.G. are named as inventors. **Data and materials availability:** Processed data not provided in the supplementary data files are available through our TCGA Publication Page (https://gdc.cancer.gov/about-data/publications/ATACseq-AWG). This includes pan-cancer raw and normalized counts matrices, cancer type–specific peak calls, cancer type–specific raw and normalized count matrices, and bigWig track files for all technical replicates. Raw ATAC-seq data as fastq or aligned BAM files will be made available through the NIH Genomic Data Commons portal (https://portal.gdc.cancer.gov/). ATAC-seq data corresponding to human plasmacytoid dendritic cells and myeloid dendritic cells (the only non-TCGA data generated here) is available through SRA BioProject PRJNA491478. The ATAC-seq peak accessibility and computed peak-to-gene linkage predictions are publicly available for interactive visualization and exploration at the UCSC Xena Browser (https://atacseq.xenahubs.net). Sample-level ATAC-seq data across all 404 donors assayed can be visualized side-by-side with all other data from TCGA, including gene expression, DNA methylation from both Illumina 450K array and WGBS platforms, and ELMER enhancer analysis results, as well as the latest survival data and mutation calls from the Genomic Data Commons. ATAC-seq data can be queried by gene, genomic position, or individual peaks. The UCSC Xena Browser makes this rich resource available for interactive online analysis and visualization by the larger scientific community. Samples from the TCGA project can only be used for TCGA efforts owing to restrictions in the material transfer agreement used for acquisition. No external groups can access the tissue or analytes.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6413/eaav1898/suppl/DC1
TCGA Analysis Network Collaborators
Materials and Methods
Protocol S1
Figs. S1 to S8
References (*55*–*78*)
Data S1 to S10

22 August 2018; accepted 28 September 2018
10.1126/science.aav1898

RESEARCH ARTICLE SUMMARY

LIMB REGENERATION

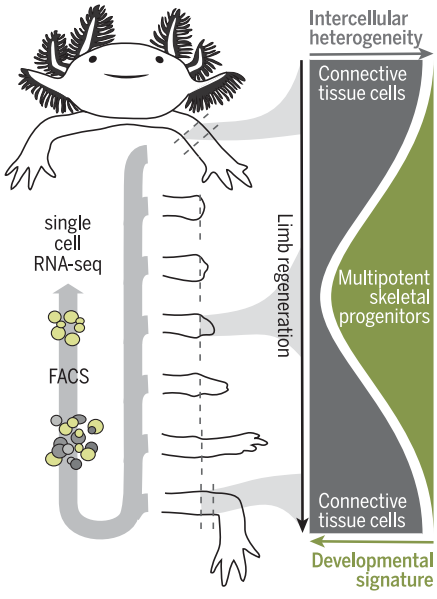
Single-cell analysis uncovers convergence of cell identities during axolotl limb regeneration

Tobias Gerber*, Prayag Murawala*†, Dunja Knapp*, Wouter Masselink, Maritta Schuez, Sarah Hermann, Malgorzata Gac-Santel, Sergej Nowoshilow, Jorge Kageyama, Shahryar Khattak, Joshua D. Currie, J. Gray Camp, Elly M. Tanaka†, Barbara Treutlein†

INTRODUCTION: Axolotls (*Ambystoma mexicanum*) and other salamanders are the only tetrapods that can regenerate whole limbs. During this complex process, changes in gene expression regulate the outgrowth of a new appendage, but how injury induces limb cells to form regenerative progenitors that differentiate into diverse cell fates is poorly understood. Tracking and molecular profiling of individual cells during limb regeneration would resolve distinct differentiation pathways and provide clues for how cells convert from a mature resting state into regenerative cell lineages.

RATIONALE: Axolotl limbs are composed of many different cell types originating from neural, myogenic, epidermal, and connective tissue (CT) lineages. Upon limb amputation, cells from nearby the amputation plane accumulate in a distinctive tissue called the blastema, which serves as a progenitor cell source to build the new limb. Transgenic axolotl strains in which descendants of distinct adult cell types can be labeled, tracked, and isolated during the regenerative process provide an opportunity to understand how particular cell lineages progress during blastema formation and subsequent limb regrowth. Combining transgenic axolotl strains with single-cell RNA sequencing (scRNA-seq) enables the tracking of individual cell types, as well as the reconstruction of the molecular steps underlying the regeneration process for these particular cell lineages. CT cells, descendants of lateral plate mesoderm, are the most abundant lineage contributing to the blastema and encompass bone and cartilage, tendons, periskeleton, and dermal and interstitial fibroblasts. These cells detect the position of the amputation site, leading to the regeneration of appropriate limb parts and making the CT a key cell lineage for deciphering and understanding molecular programs of regeneration.

RESULTS: We used an inducible Cre-loxP fluorescence system to establish genetically marked transgenic axolotl strains for isolating CT cells from adult limb tissue as well as CT descendants in the blastema. We used scRNA-seq to molecularly profile CT cells along a dense time course of blastema formation and the outgrowth of the regenerated arm, as well as stages of embryonic limb development. This profiling indicated that CT cells express adult phenotypes that are lost upon the induction of regeneration. The heterogeneous population of CT-derived cells converges into a homogeneous and transient blastema progenitor state that at later stages recapitulates an embryonic limb bud–like program. Notably, we did not find evidence of CT stem cells or blastema-like precursors in the mature arm. We found that CT subtypes have spatially restricted contributions to proximal and distal compartments in the regenerated limb. Specifically, a particular CT subtype—periskeletal cells—extended the severed skeleton at the amputation site whereas fibroblastic CT cells de novo regenerated distal skeletal segments. By using high-throughput single-cell transcriptomics and Brainbow axolotl-based clonal lineage tracing, we could follow the redifferentiation trajectories of CT lineages during the final stages of regeneration. These findings established the formation of a multipotent skeletal progenitor cell that contributed to tendons, ligaments, skeleton, periskeleton, and fibroblasts.



Regeneration of the axolotl upper arm CT. Transgenic axolotl strains were used to isolate CT cells at different stages during limb regeneration. Single-cell transcriptomes were generated, and transcriptional signatures were used to track the fate of cells during regeneration. We found that heterogeneous CT cell types in the adult upper arm funnel into a homogeneous multipotent skeletal progenitor state that recapitulates axolotl limb development, before their redifferentiation into the heterogeneous CT cell subtypes.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aaq0681

CONCLUSION: CT cells are a key cell type for understanding regeneration because they form the patterned limb skeleton that guides the regeneration of the other limb tissues, such as muscle. Because of the cells' heterogeneity and intermingling with other cell types, it had been difficult to study how CT forms regenerative blastema cells. The use of these newly generated transgenic reporter strains combined with single-cell transcriptomic analysis and clonal tracing have allowed us to determine that CT cells with diverse molecular features traverse through a distinctive molecular state as they dedifferentiate into a common, multipotent progenitor resembling an embryonic limb bud cell. In the future, it will be important to test which components of the transition state are required for the dedifferentiation process. Furthermore, this work opens the possibility to examine how regeneration-associated genes and their associated chromatin structure are regulated during this transition. Lastly, the work raises the possibility that the limited regeneration seen among mammals is due to an inability to reprogram CT to such embryonic states. ■

*These authors contributed equally to this work.
The list of author affiliations is available in the full article online.
†Corresponding author. Email: barbara_treutlein@eva.mpg.de (B.T.); elly.tanaka@imp.ac.at (E.M.T.); prayag.murawala@imp.ac.at (P.M.)
Cite this article as T. Gerber *et al.*, *Science* **362**, eaaq0681 (2018). DOI: 10.1126/science.aaq0681

RESEARCH ARTICLE

LIMB REGENERATION

Single-cell analysis uncovers convergence of cell identities during axolotl limb regeneration

Tobias Gerber[1]*, Prayag Murawala[2,3]*†‡, Dunja Knapp[3]*, Wouter Masselink[2]†, Maritta Schuez[3], Sarah Hermann[3], Malgorzata Gac-Santel[1], Sergej Nowoshilow[2,3]†, Jorge Kageyama[1], Shahryar Khattak[3], Joshua D. Currie[3]§, J. Gray Camp[1], Elly M. Tanaka[2,3]†‡, Barbara Treutlein[1,4,5]‡

Amputation of the axolotl forelimb results in the formation of a blastema, a transient tissue where progenitor cells accumulate prior to limb regeneration. However, the molecular understanding of blastema formation had previously been hampered by the inability to identify and isolate blastema precursor cells in the adult tissue. We have used a combination of Cre-loxP reporter lineage tracking and single-cell messenger RNA sequencing (scRNA-seq) to molecularly track mature connective tissue (CT) cell heterogeneity and its transition to a limb blastema state. We have uncovered a multiphasic molecular program where CT cell types found in the uninjured adult limb revert to a relatively homogenous progenitor state that recapitulates an embryonic limb bud–like phenotype including multipotency within the CT lineage. Together, our data illuminate molecular and cellular reprogramming during complex organ regeneration in a vertebrate.

Among tetrapods, only salamanders have the capacity to replace a lost limb (*1*). The adult axolotl (*Ambystoma mexicanum*) limb is composed of many different cell types originating from neural, myogenic, epidermal, and connective tissue (CT) lineages. Upon limb amputation, cells from areas near the amputation plane accumulate in a tissue called the blastema. Cells within the blastema are capable of fully regenerating the missing limb [reviewed in (*2*)]. CT cells, descendants of lateral plate mesoderm (LPM), are the most abundant lineage contributing to the blastema (*3*–*5*) and encompass bone and cartilage, tendons, periskeleton, and dermal and interstitial fibroblasts. The CT cells are a key cell type for deciphering the molecular program of regeneration, as they express factors that guide the regeneration of appropriate limb parts (*4*, *6*–*8*). Nevertheless, how mature CT produces blastema cells has not been molecularly defined because of the inability to isolate and deconstruct this cell population. It is even unclear whether mature CT cells molecularly "reprogram" into embryonic cell–like limb progenitors or whether the CT harbors preexisting stem cells that selectively seed the blastema.

We generated Cre-loxP reporter lines to track CT compartments and uncover compartment-associated contributions to different limb segments. We used single-cell mRNA sequencing (scRNA-seq) to dissect CT heterogeneity in the blastema, as well as the adult, embryonic, and regenerated forelimb, enabling the characterization of CT cell types and lineage relationships as cells regenerated the limb. Notably, CT cells lose their mature phenotypes to form multipotent limb bud–like progenitors in the blastema. The combination of lineage tracking and single-cell transcriptomics resolves the origin and molecular profiles of redifferentiated CT cell types that emerge from the blastema. Our work provides a molecular view of individual cells that build a blastema and reconstitute a patterned limb skeleton.

A Cre-loxP reporter system tracks CT cells during axolotl limb regeneration

To label and track axolotl CT cells, we generated a germline transgenic line that expresses the tamoxifen-inducible Cre-ERT2 (Cre-ER) gene under the control of the *Prrx1* limb enhancer element (*Prrx1:TFPnls-T2A-Cre-ERT2;CAGGs:lp-GFP-3pA-lp-Cherry*, hereinafter referred to as *Prrx1:Cre-ER;Caggs:lp-Cherry* animals) (figs. S1 and S2 and tables S1 and S2). *Prrx1*, a paired-related homeobox gene, is expressed in CT precursors in the developing limb bud and in the limb blastema (*9*–*11*). Immunofluorescence staining of PRRX1 protein confirmed specific expression in axolotl limb bud cells (Fig. 1A), blastema cells (Fig. 1B), and adult CT (fig. S2). Administration of tamoxifen to *Prrx1:Cre-ER;Caggs:lp-Cherry* animals at the limb bud stage resulted in an efficient (>80%) genetic labeling of adult limb CT (Fig. 1, C and D, and fig. S1E). Notably, after limb amputation, we found that *Prrx1*-expressing blastema cells express mCherry, showing that the transgenic reporter efficiently marks the adult precursors to the blastema cells (Fig. 1B). Examination of regenerates 25 days postamputation (dpa) revealed mCherry-expressing cells in upper and lower arm CT (Fig. 1D and fig. S1, C to F), showing that CT gives rise to new CT during regeneration. Therefore, this transgenic line provides a system to track CT cells during limb regeneration.

We used a high-throughput droplet-based scRNA-seq method (10× Genomics) to sample the cellular diversity in the uninjured adult limb and further validate this transgenic line. We converted cells at the limb bud stage and performed scRNA-seq on the dissociated uninjured adult limb tissue containing labeled and unlabeled cells (2379 cells) (table S3). Using unbiased clustering and the expression of marker genes, we identified endothelial, epidermal, immune, muscle, red blood, and CT cells (Fig. 1E). mCherry mRNA from *Prrx1:Cre-ER;Caggs:lp-Cherry* converted cells was detected only in the CT cluster, which included periskeletal, tendon, dermal, and fibroblastic cell subpopulations as identified on the basis of canonical marker expression (Fig. 1F). To specifically examine CT heterogeneity, we used *t*-distributed stochastic neighbor embedding (tSNE) clustering to analyze 2375 single-cell transcriptomes after the isolation of labeled *Prrx1:Cre-ER;Caggs:lp-Cherry* line–derived CT cells by fluorescence-activated cell sorting (FACS) from the adult upper forelimb (Fig. 2, A and B, and table S5). We identified eight distinct clusters that we assigned on the basis of marker gene expression as tenocytes (identified by the expression of marker gene *Tnmd*), periskeletal cells (with marker gene *Col8a2*), actively cycling cells (with marker gene *Ccnb1*), and five fibroblastic CT subpopulations (fCT I to V) (Fig. 2, B and C, and table S6). The two fibroblastic CT populations fCT IV and fCT V could be identified as components of the dermis on the basis of *Twist2* and *Ptgds* expression. The cycling population contained multiple different cell types, including tenocytes, fCT I, fCT III, and fCT IV (fig. S4A). We detected only very few skeletal cells (fig. S4B), likely because of their inability to dissociate from the skeletal matrix. Because of the low cell number, skeletal cells fall within the periskeletal cluster in the tSNE plot. Previous live imaging data (*12*) and our own cell tracking data (fig. S9) show that skeletal cells do not contribute to the regeneration. Since regeneration of bone progresses via endochondral

[1]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany. [2]Research Institute of Molecular Pathology (IMP), Vienna BioCenter (VBC), Campus-Vienna-Biocenter 1, 1030 Vienna, Austria. [3]Deutsche Forschungsgemeinschaft (DFG) Center for Regenerative Therapies (CRTD), Technische Universität Dresden, Fetscherstraße 105, 01307 Dresden, Germany. [4]Max Planck Institute of Molecular Cell Biology and Genetics, 108 Pfotenhauerstraße, 01307 Dresden, Germany. [5]Department of Biosciences, Technical University Munich, 85354 Freising, Germany.
*These authors contributed equally to this work.
†Present address: IMP, Campus-Vienna-Biocenter 1, 1030 Vienna, Austria.
‡Corresponding author. Email: barbara_treutlein@eva.mpg.de (B.T.); elly.tanaka@imp.ac.at (E.M.T.); prayag.murawala@imp.ac.at (P.M.)
§Present address: Department of Cell and Systems Biology, University of Toronto, 25 Harbord St., M5S 3G5 Toronto, Canada.

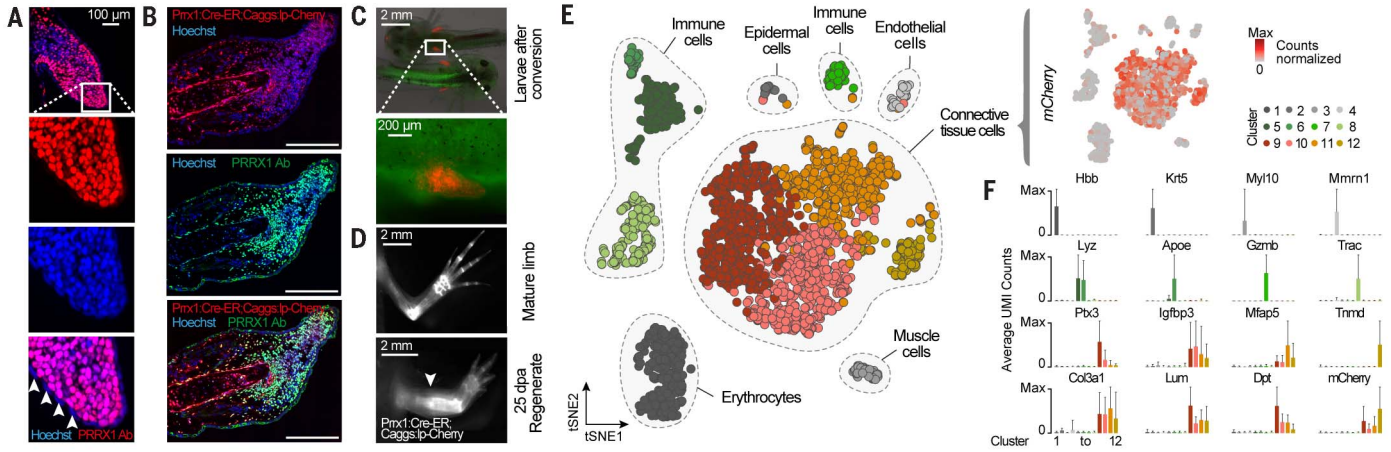


Fig. 1. Tracking and molecular profiling of axolotl limb CT. (**A**) Longitudinal section of a stage 47 limb bud stained with anti-PRRX1 antibody (Ab) (red) identifying PRRX1 as a pan-CT marker during limb development. Arrowheads indicate the absence of PRRX1 staining in the epidermis. The boxed area in the top panel corresponds with the higher-magnification views in the bottom three panels. (**B**) Longitudinal section of a blastema 11 dpa, stained with anti-PRRX1 Ab (green). Red, converted cells; blue, Hoechst-stained nuclei. Scale bars, 500 μm. (**C**) Embryos after induction of *Prrx1:Cre-ER;CAGGs:lp-Cherry* by 4-OHT showing expression of mCherry only in limb mesenchyme. (**D**) Fluorescence image of converted cells in an uninjured limb and a regenerated limb (conversion at limb bud stage) indicating stable labeling of CT before and after regeneration. The arrowhead indicates the amputation plane. (**E**) (Left) tSNE plot visualizing scRNA-seq data for 2379 single cells (circles) from the adult axolotl upper arm. Dashed lines outline related cell types. (Right) mCherry expression is detected exclusively in CT cell types. (**F**) Bar plots showing the mean expression of marker genes in each cluster. The *x* axis represents cell clusters identified in (E). Error bars indicate SD. UMI, unique molecular identifier.

ossification, "skeletal cells" here refers to osteogenic and chondrogenic cells, the proportions of which differ between mature and regenerating bone. In summary, these profiling data validate the specificity of our *Prrx1:Cre-ER;Caggs:lp-Cherry* reporter animals, provide a cell atlas and a marker set for cell types of the uninjured adult axolotl limb (table S4), and characterize the heterogeneity of the upper arm CT (table S6).

Blastema formation from axolotl upper arm CT cells involves molecular funneling during regeneration

To understand the molecular pathways involved in CT regeneration, we used a high-transcriptome-coverage scRNA-seq method (Fluidigm C1) to analyze mCherry-positive (mCherry$^+$) cells isolated by FACS (*Prrx1:Cre-ER;Caggs:lp-Cherry*) along a dense time course that captures the major transitions during regeneration. Time points were determined on the basis of the average blastema cell cycle length (53 to 103 hours) (*13*, *14*) and previous bulk transcriptional dynamics (*15*) and therefore included the uninjured upper forelimbs (108 cells), as well as blastema stages (3 dpa, 108 cells; 5 dpa, 167 cells; 8 dpa, 121 cells; 11 dpa, 163 cells; and 18 dpa, 135 cells) and a fully regenerated upper forelimb (3 to 12 months postamputation, 128 cells) (Fig. 2A, fig. S3, and table S7). We first analyzed the uninjured upper forelimb data and could confirm the presence of the CT subpopulations described above, which allowed identification of additional marker genes for each cell type (fig. S4, C and D). We next combined fibroblastic and periskeletal cells from the uninjured CT, which are the major populations contributing to the blastema (*12*), with all blastema states and the regenerated CT and used a diffusion pseudotime estimate (*16*, *17*) to reconstruct the lineage path from uninjured CT through the blastema state to the regenerated CT (Fig. 2D and fig. S4E). This analysis revealed a general time-dependent progression, with some intermixing between time points. Diffusion component 3 visualizes how heterogeneous differentiated CT cells funnel into the relatively homogeneous early blastema state (Fig. 2E) and regenerated CT cells funnel out of the blastema to reestablish the initial cellular heterogeneity (fig. S4F). We confirmed this observation with two alternative analyses. When focusing on the cell type–specific expression patterns found in the uninjured tissue (Fig. 2F), we did not observe a comparable heterogeneity within the blastema cell populations and instead found that heterogeneity was diminished at all blastema time points (Fig. 2G) up until the redifferentiation of CT cell types (fig. S4G). Further, we calculated for each time point the number of unique genes detected within the top 100 most expressed genes per cell as a measure of intercellular heterogeneity per time point. We detected a decrease in the number of unique genes from the adult uninjured upper limb through the early blastema stages up to the 11-dpa blastema, followed by an increase at 18 dpa and in the regenerate (fig. S4H). This is consistent with a progressive decrease in intercellular heterogeneity during blastema formation until 11 dpa, with the highest heterogeneity found in the uninjured adult upper arm, the regenerate, and the 18-dpa blastema when cells begin to differentiate. Notably, we did not find blastemalike cells within the uninjured tissue because all cells from the adult upper forelimb cluster separately from blastema-derived cells in a tSNE analysis (fig. S4I). Also, when the cycling cells in the uninjured tissue are analyzed together with cycling blastema cells, the cells derived from the uninjured tissue are distinct because they express signatures of differentiated CT cell types (e.g., *Mfap5* and *Fbn1*) and lack expression of blastema markers (e.g., *Nrep* and *Prdx2*) (fig. S4, J and K).

Our molecular data taken together with previous cell lineage observations (*12*) and digital tracking data (*18*) show little numerical bias in cell contribution during blastema formation, allowing us to conclude that blastema formation does not involve the selection of a preexisting stem cell population. With the cell cycle length (*13*, *14*) and cell counts pointing to an expansion of cell numbers by an average of six times from the uninjured to the 11-dpa time point (supplementary materials and methods), a "preexisting" blastema cell would need to constitute at least 20% of the mature CT population to account for the cell expansion. In summary, these data indicate that mature CT cells dedifferentiate into a relatively homogeneous pool of progenitor cells when forming the blastema.

Next, we explored the transcriptional changes that ensue during the regenerative process within the blastema (Fig. 2H and fig. S5, A to C). The uninjured CT state is characterized by the expression of many extracellular matrix (ECM) genes that are down-regulated during blastema formation (Fig. 2, H and I). We observed an inflammation response (e.g., expression of *Il11* and *Cxcl2*) in the early blastema (3 to 5 dpa) coinciding with the expression of various matrix metalloproteinase genes (*Mmp8* and *Mmp13*),

indicating extensive ECM remodeling induced by injury. By 8 dpa, cells express genes involved in proliferation (e.g., *Ccnb3*), and by 11 dpa, there is positive regulation of transcription (*Hmga2* and *Lbh* expression) and evidence of repatterning (*Hoxa13* and *Hoxd12* expression). Reestablishment of ECM (via *Matn4*) is initiated by 11 dpa but peaks at 18 dpa. Notably, genes necessary for skeletal development (*Hoxd13*, *Col2a1*, and *Sox9*) are detected in the blastema 18 dpa. In addition, the expression of genes for various signaling molecules (*Dkk1*, *Bambi*, and *Lep*) associated with regeneration are induced throughout blastema development (*15*, *19*, *20*). These data constitute a molecular map of CT cells as they transition through the limb-regenerative process.

CT reprogramming progresses through a blastema-specific state before recapitulating embryonic limb bud development

We next explored the extent to which CT regeneration recapitulates limb development. To cover different stages of limb development, we examined the single-cell transcriptomes of stage 28 limb field cells (82 cells), as well as stage 40 (76 cells) and stage 44 (121 cells) limb bud cells (*21*) (Fig. 3A and table S7) and identified CT cells on the basis of *Prrx1* expression (279 cells). By stage 44, the limb bud contains approximately 1500 cells. We observed that CT precursors in the developing limb are distinct from and less differentiated than adult CT (Fig. 3B and fig. S6A). Although cells from the two limb bud stages showed major similarities, cells from the stage 28 limb field were distinct from stage 40 and 44 limb bud cells, as shown by the unique expression of genes such as *Wnt8a* or *Hoxd1* (fig. S6B). Variation in the expression of spatial patterning genes constituted the major source of heterogeneity in the limb buds (stages 40 and 44) (fig. S6C), whereas patterning genes were generally not yet expressed in stage 28 limb field cells (Fig. 3C). This variation enabled reconstruction of the proximal-distal and anterior-posterior axes through the correlated expression of patterning genes, confirming that our data captured the representation of the limb bud cells (Fig. 3D) (*22*). We investigated when patterning

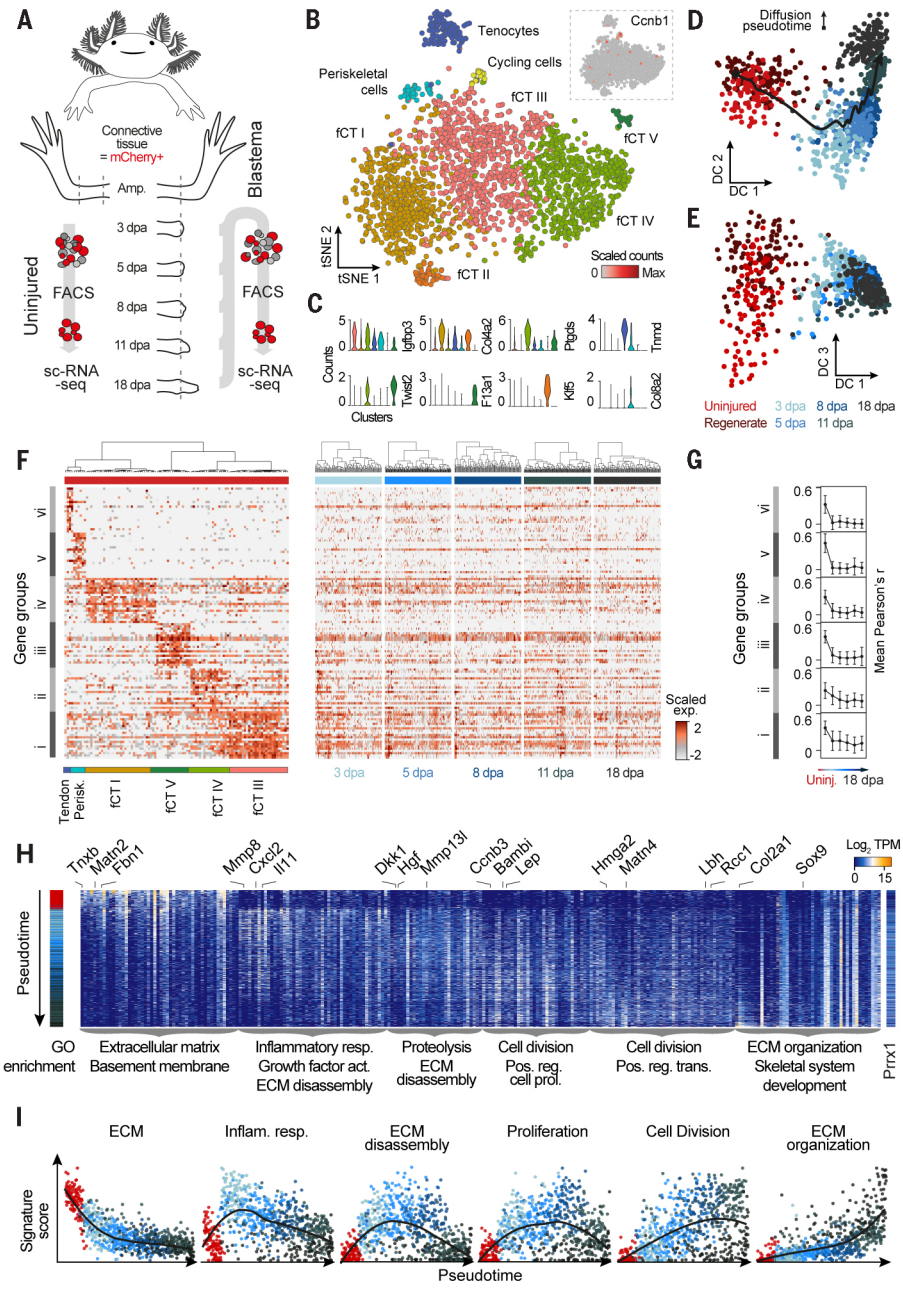


Fig. 2. Blastema formation from axolotl upper arm CT cells involves molecular funneling during regeneration. (**A**) Schematic of CT scRNA-seq experiments performed on mCherry^{+} CT cells of the uninjured axolotl upper arm (0 dpa) and the regenerating upper arm blastema (3, 5, 8, 11, and 18 dpa) from *Prrx1:Cre-ER;Caggs:Lp-Cherry* animals (conversion at 1-cm size). Cells were sorted by FACS. Amp., amputation. (**B**) A tSNE projection of 2375 single-cell transcriptomes visualizes the cellular heterogeneity of the uninjured upper arm CT and reveals eight clusters that refer to seven CT subtypes and one cluster of cycling cells (with expression of *Ccnb1*) (inset). (**C**) Violin plots showing distribution of expression for selected tSNE cluster marker genes (B). Colors refer to tSNE clusters (B). (**D**) Diffusion map projection (*17*) describing lineage relationships between uninjured CT cells, blastema-derived cells, and cells from a regenerated upper arm. DC, diffusion component. (**E**) DC 3 captures the cell type heterogeneity in the uninjured CT that is lost in the blastema. (**F**) Expression of cell type marker genes (groups i to vi) identified for the uninjured CT, shown as a heat map for uninjured CT and blastema time points, with genes in rows and cells hierarchically clustered in columns. Transcript levels are scaled across rows. Perisk., periskeleton. (**G**) Mean pairwise correlation (Pearson) between genes of each gene group (F) across all cells, calculated for each time point. Mean correlation coefficients decrease over the course of blastema formation. Error bars, SD. (**H**) Heatmap visualization of time point–specific marker genes (columns) with cells (rows) ordered by diffusion pseudotime (see also fig. S4I). GO enrichments (bottom) and exemplary genes (top) are shown (see also fig. S5A). Colored sidebar, time points. resp., response; act., activity; pos. reg., positive regulation; prol., proliferation; trans., transcription. (**I**) Pseudotemporal expression of different gene signatures across all cells from uninjured upper arm CT to 18-dpa blastema. Conditional means smoothed by using LOESS (local regression) are presented.

axes start to be established during limb regeneration and found that anterior-posterior markers (*Fgf8* and *Shh*, respectively) and proximal-distal markers [*Meis2* (proximal) and *Hoxa11* and *Hoxa13* (distal)] are established between 8 and 11 dpa (Fig. 3E). In a manner similar to that for the limb bud, patterning genes enabled the spatial reconstruction of cell positions in a blastema 11 dpa (Fig. 3F and fig. S6D).

We next quantified the similarity between different blastema stages and limb development. We created mock bulk transcriptomes for each uninjured, regenerate, embryonic, and blastema time point and calculated the correlation (Spearman's *r*) of each single cell with each bulk transcriptome. We found that the correlation of blastema cells with stage 40 and stage 44 limb buds peaks at 11 dpa, with a progressively weaker correlation at earlier blastema time points (Fig. 3G and fig. S6, E and F). Notably, we observed the same trend when we compared blastema cells with the stage 28 limb field (Fig. 3H). Additionally, we found that cells in the 3- and 5-dpa blastemas are similar neither to uninjured adult cells nor to limb bud or limb field cells, suggesting that a cell state emerges in the blastema that is distinctive for limb regeneration (Fig. 3I and fig. S6E). Within our dataset, we identified many genes that were expressed only in the early blastema stages, whereas 11-dpa cells largely share expression patterns with the cells derived from the developing limb (Fig. 3J and fig. S6, G and H). Our data suggest that the former CT cells within the blastema initiate reprogramming by using genetic mechanisms that are distinct from embryonic limb bud development but arrive at a limb bud–like state by day 11.

Tracking different CT subpopulations in the blastema reveals distinct cell sources for proximal versus distal limb regenerate tissue

We next sought to analyze how different CT cell subpopulations contribute to redifferentiated cell types in the regenerated limb. We first generated a new transgenic line by using the *Col1a2* promoter (*Col1A2:TFPnls-T2a-ERT2-Cre-ERT2;CAGGs:lp-GFP-3pA-lp-Cherry*, hereinafter referred to as *Col1A2:ER-Cre-ER;Caggs:lp-Cherry*) in which only a subset of CT cells are labeled and tracked (fig. S7). Specifically, conversion of 3-cm larvae leads to limbs with genetic labeling primarily of the skeleton, periskeleton, and tendons but very few dermal fibroblasts and no interstitial fibroblasts, thus constituting a subset of cell types labeled in the *Prrx1:Cre-ER;Caggs:lp-Cherry* line (Fig. 4A and fig. S7A). scRNA-seq on 36 sorted, labeled cells from the uninjured adult upper arm revealed periskeletal cells and tenocytes (bone cells were not recovered in the dissociation, but they do not contribute to the blastema), confirming the histological analysis of the labeled cells (Fig. 4B and table S8). We next performed scRNA-seq on labeled *Col1a2:ER-Cre-ER;Caggs:lp-Cherry* (referred to as *Col1a2*) descendants in the 11-dpa blastema (349 cells) and could identify undifferentiated progenitors as well as immature skeletal and nonskeletal cell lineages (Fig. 4C and table S8). A pseudotemporal trajectory displayed two branching paths: one transitioning from progenitors (expressing *Mycl*, *Matn4*, and *Nrep*) to nonskeletal cells (expressing *Tnmd*, *Fcn2*,

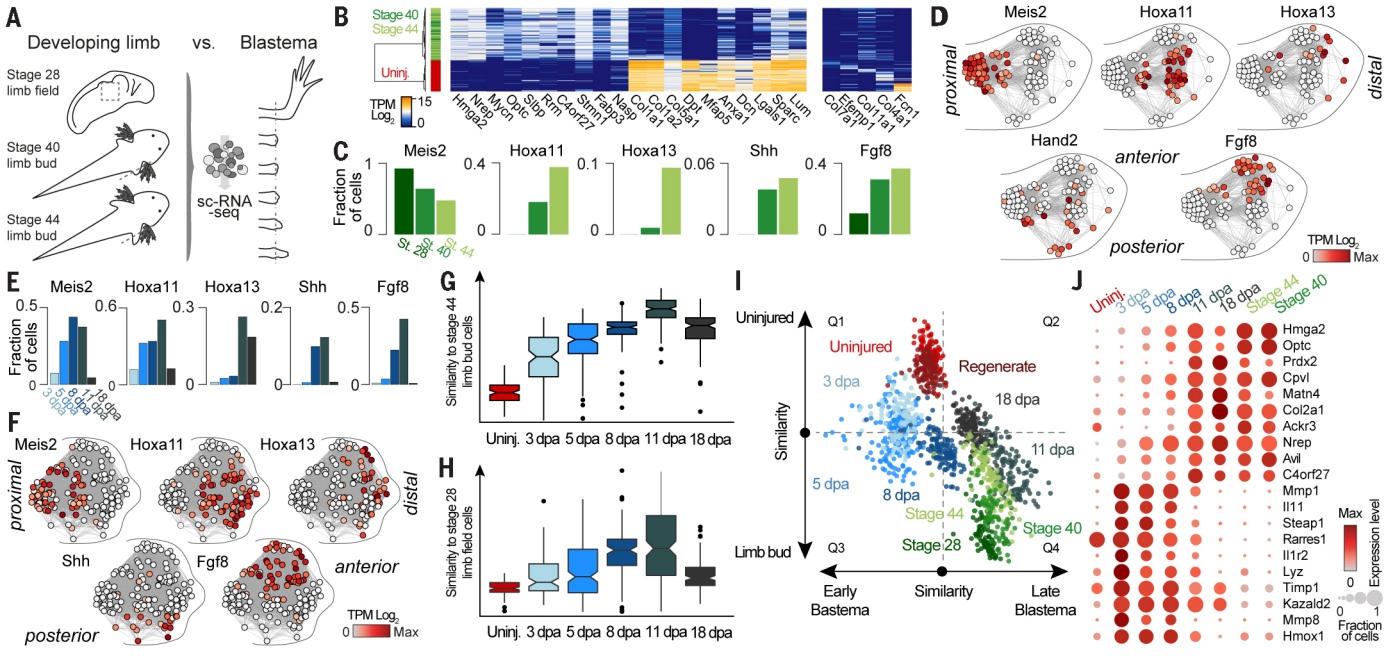


Fig. 3. CT reprogramming progresses through a blastema-specific state before recapitulating embryonic limb bud development. (**A**) Overview of scRNA-seq experiments on three axolotl limb bud stages. Limb bud CT cells (279) were in silico identified based on *Prrx1* expression and compared with blastema cells on the transcriptome level. (**B**) Heat map showing expression of genes (columns) that distinguish mature limb CT cells from limb bud CT cells (rows, hierarchically clustered). The expression of CT cell type marker genes is shown on the right. (**C**) Bar graphs showing the fraction of cells per embryonic stage that express proximal-distal (*Meis2*, *Hoxa11*, and *Hoxa13*) or anterior-posterior (*Fgf8* and *Shh*) patterning genes. (**D**) Intercellular correlation network of stage 44 limb bud cells (circles) placing cells in a hypothetical position within an imaginary limb bud on the basis of the expression of five known patterning genes (see also fig. S6C). (**E**) Limb bud patterning genes are reactivated during blastema formation. Bar graphs show the fraction of cells per blastema time point that express proximal-distal (*Meis2*, *Hoxa11*, and *Hoxa13*) or anterior-posterior (*Fgf8* and *Shh*) patterning genes. (**F**) Intercellular correlation network of 11-dpa blastema cells (circles) placing cells in a hypothetical position within an imaginary limb blastema on the basis of the expression of five patterning genes (see also fig. S6D). (**G** and **H**) Box plots showing distributions of scaled correlation between single-cell transcriptomes at any given time point and the mock bulk transcriptome of stage 44 limb bud (G) or stage 28 limb field (H) CT cells. Limb bud and limb field progenitors are most similar to 11-dpa blastema cells. (**I**) Scatter plot showing differential correlation of single-cell transcriptomes (dots; color represents time point) with limb bud versus mature CT transcriptomes (*y* axis) and with 5-dpa blastema versus 11-dpa blastema transcriptomes (*x* axis). (**J**) Dot plot visualizing the expression of genes shared between the 11-dpa blastema and limb bud progenitor cells. The circle size represents the fraction of cells at each time point expressing the gene, and color represents the average expression level.

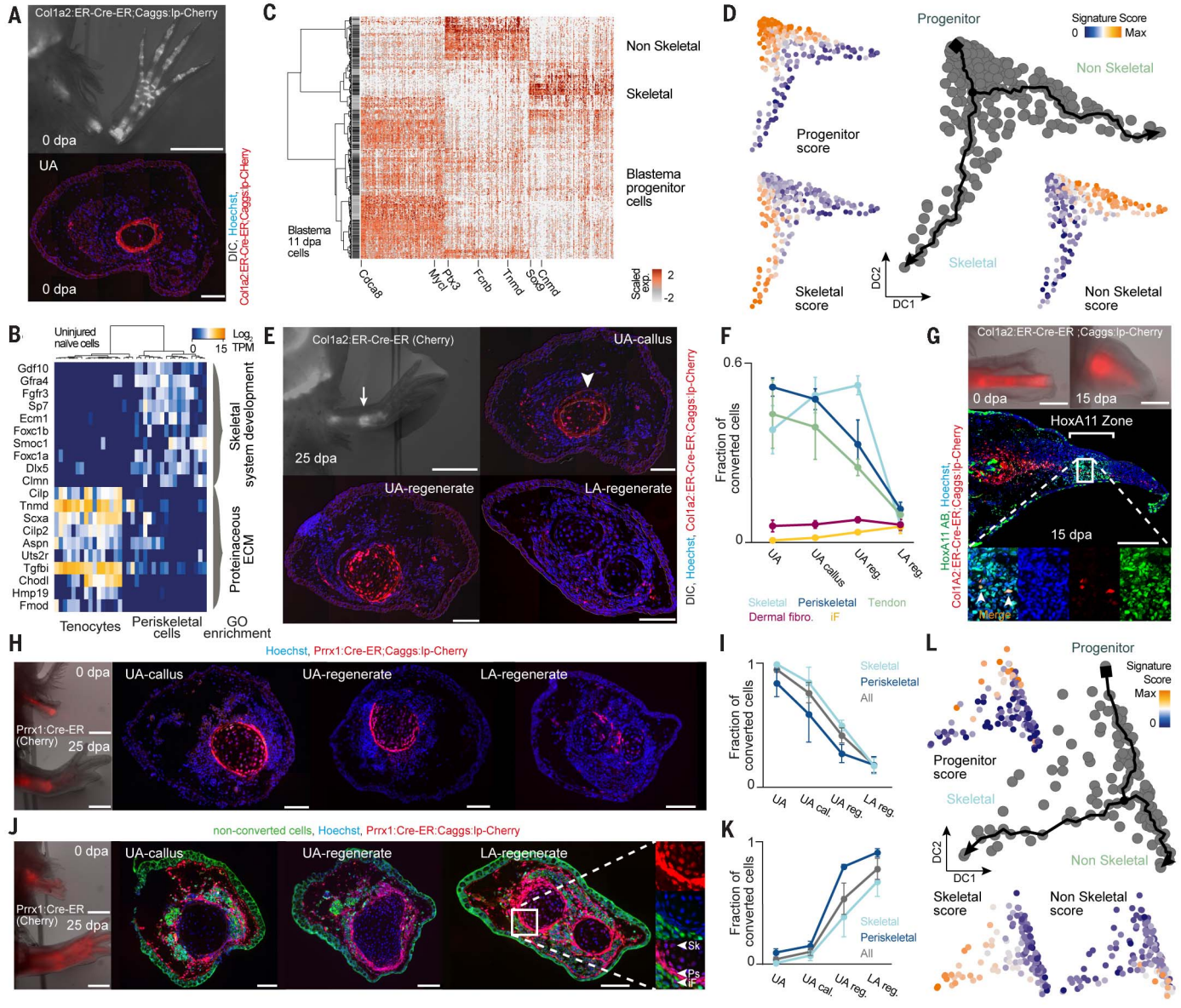


Fig. 4. Tracking CT subpopulations in the blastema reveals distinct cell sources for proximal versus distal limb regenerate tissue. (**A** to **G**) *Col1a2:ER-Cre-ER;Caggs:lp-Cherry*–labeled animals were used. (A) (Top) Bright-field image with fluorescent overlay showing limbs at 0 dpa. (Bottom) Upper arm (UA) limb cross section revealing periskeletal and tendon cell labeling (Hoechst, blue; *mCherry*, red). Scale bars, 2 mm. (B) Heat map showing marker expression for 36 mature limb periskeletal and tendon cells (hierarchically clustered; Pearson). (C) Heat map showing scaled expression (exp.) of genes identified by PCA (table S10) in 349 Col1a2 line–derived 11-dpa blastema cells. (D) Diffusion map projection of Col1a2 line–derived 11-dpa blastema cells with signature scores shown. (E) (Top left) Image overlay of 25-dpa limbs (scale bar, 2 mm). The arrow shows the site of amputation. (Top right, bottom left, and bottom right) Limb cross sections along the proximal-distal axis. Samples are stained with Hoechst (blue), and images are overlaid with DIC and the mCherry fluorescence signal (red) of converted cells. The arrowhead indicates callus formation. Scale bars, 200 μm. (F) Fraction of converted cells in five CT subtypes at different proximal-distal positions after regeneration (n > 8000 cells, three limbs). reg., regenerate; fibro., fibroblasts; iF, interstitial fibroblasts (fCT I to III). (G) *mCherry*-labeled humerus transplantation into an unlabeled host. (Top) Images of live animals. Scale bars, 1 mm. (Center) Longitudinal sections of 15-dpa blastema showing converted cells (red), Hoxa11 expression (green), and Hoechst staining (blue). Scale bar, 500 μm. (Bottom) Colocalization (arrowheads) of mCherry$^+$ and Hoxa11$^+$ cells. (**H** to **L**) *Prrx1:Cre-ER;Caggs:lp-Cherry*–labeled animals were used. (H) *mCherry*-labeled humerus transplantation into an unlabeled host. (Left to right) Images of live animals (scale bars, 1 mm) and limb cross sections from UA-callus, UA-regenerate, and LA-regenerate positions (15 dpa; converted cells, red; Hoechst, blue; scale bars, 200 μm). (I) Quantification of converted cells among periskeletal and skeletal cells and an aggregate of both subtypes (All) along the proximal-distal axis. cal., callus. (J) Unlabeled humerus transplantation into an *mCherry* converted host. (Left to right) Images of live animals (scale bars, 1 mm) and limb cross sections from UA-callus, UA-regenerate, and LA-regenerate positions (15 dpa; converted cells, red; unconverted cells, green; Hoechst, blue; scale bars, 200 μm). The magnification shows mCherry$^+$ periskeletal (Ps) and skeletal (Sk) cells and interstitial fibroblasts (iF) (fCT I to III). (K) Quantification of converted cells among periskeletal and skeletal cells and an aggregate of both subtypes along the proximal-distal axis. Error bars in (F), (I), and (K) indicate SD. (L) Diffusion map projection of *Prrx1* line–derived 18-dpa blastema cells with signature scores shown. The map was created by using genes identified by PCA (table S10).

and *Aspn*; likely precursors to tenocytes and periskeletal cells) and the other to skeletal cells (expressing *Cnmd*/*Lect1* and *Otos*) (Fig. 4D, fig. S6I, and table S10).

Microscopic examination of regenerates revealed a spatial bias in cell contribution. Labeled cells were found primarily in the extension to the existing bone at the amputation site (humerus), with few cells in more distal segments (Fig. 4, E and F). We also found a similar spatial restriction of *Col1a2* descendants to extending bone in lower arm amputations (fig. S8B). This is in contrast to the *Prrx1:Cre-ER;Caggs:lp-Cherry* line in which all proximal and distal positions of the regenerate are labeled. These data provide direct evidence supporting a previous hypothesis that distinct cell sources are used for bone extension versus de novo segment formation (*23*). The *Col1a2* line–derived blastema cells could be intrinsically or extrinsically limited to extending existing bone. We transplanted a humerus from a *Col1a2:ER-Cre-ER;Caggs:lp-Cherry* donor into an unlabeled host and found that, after upper arm amputation, some rare mCherry$^+$ cells were found in the distal, *Hoxa11*$^+$ region of the blastema 15 dpa and showed Hoxa11 staining (Fig. 4G). These observations lead us to conclude that the *Col1a2* line–derived blastema cells are not intrinsically limited in their segmental identity but rather are strongly associated with the callus and therefore spatially biased toward extending their bone of origin.

We performed multiple transplantation experiments to confirm the above-mentioned observations and resolve the source of the distal CT. To exclude the possibility that the *Col1a2* driver marked only a subset of periskeleton and tendon with spatially restricted potential, we grafted a humerus from *Prrx1:Cre-ER;Caggs:lp-Cherry* (referred to as *Prrx1*) converted limbs into an unlabeled host limb, replacing the host humerus before upper arm amputation (Fig. 4, H and I). As in the *Col1a2* tracing, mCherry$^+$ cells were found in the callus and the regenerated periskeleton and skeleton just beyond the amputation site, with progressive depletion toward the lower arm regenerate. We next performed a complementary graft of unlabeled humerus into *Prrx1* converted hosts to label nonbone CT. We found nearly complete labeling of tendon, skeleton, and periskeleton as well as dermal and interstitial fibroblasts in the lower arm regenerate, indicating that nonskeletal CT cells regenerate the distal CT (Fig. 4, J and K). We separately grafted a muscle fiber bundle containing labeled interstitial fibroblasts from embryonic *Prrx1* converted animals into unlabeled hosts and found extensive labeling of the distal CT (fig. S10). These data indicated that the fibroblastic CT cells were the major contributors to the distal limb regenerate. To analyze the regeneration of distal limb CT on the molecular level, we performed a lineage reconstruction analysis on the scRNA-seq data of labeled *Prrx1* descendants in the 18-dpa blastema (Fig. 4L, fig. S6J, and table S10). This analysis revealed a bifurcated path where uncommitted blastema progenitors branch

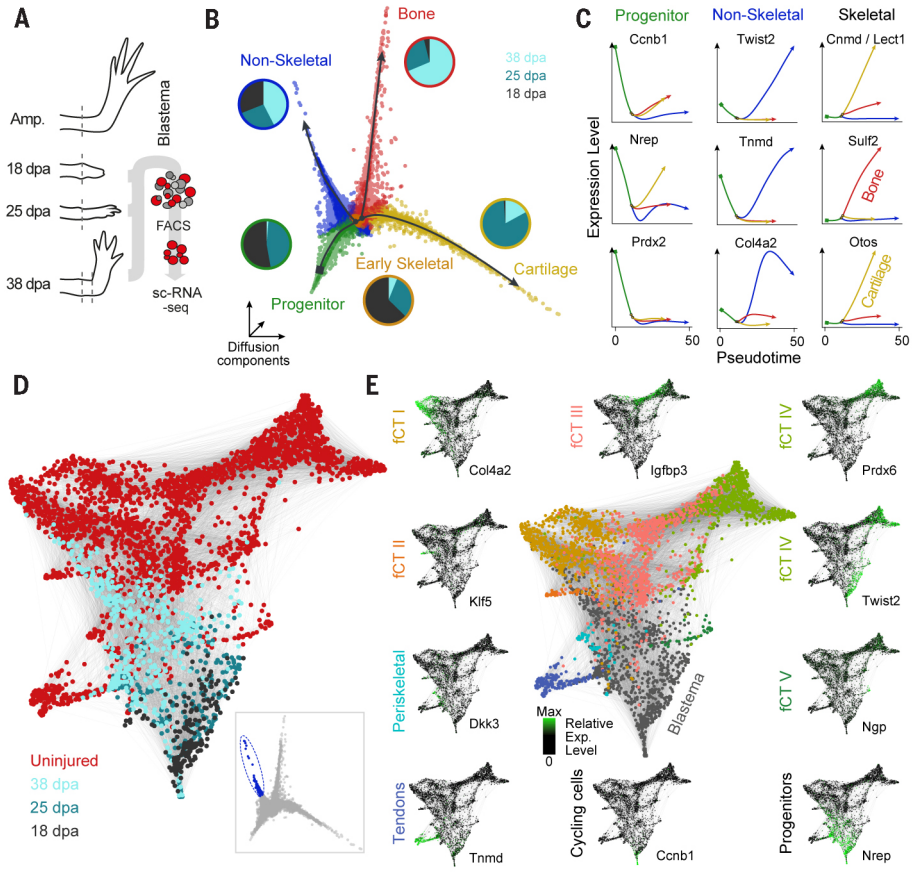


Fig. 5. CT lineages reemerge through multipotent progenitors. (**A**) High-throughput scRNA-seq was performed on mCherry$^+$ CT cells of the uninjured axolotl upper arm (0 dpa) and the upper arm blastema (at 18, 25, and 35 dpa) from *Prrx1:Cre-ER;Caggs:Lp-Cherry* converted animals. Cells were sorted by FACS. (**B**) Diffusion analysis (*17*) and tree construction (*24*) of blastema time points (18, 25, and 38 dpa) identify four main branches. Pie charts show time point proportions per branch. (**C**) Pseudotemporal marker gene expression for each branch. (**D**) SPRING analysis (*48*) of the nonskeletal blastema branch cells (blue inset) together with mature CT cells (a total of 3151 cells) reveals the reemergence of CT subpopulations identified in the mature tissue. (**E**) SPRING plots colored by early blastema (*Nrep*), cell cycle (*Ccnb1*), and CT subtype (*Tnmd*, *Col4a2*, and *Twist2*) marker expression show the cell differentiation during late phases of regeneration.

off into a nonskeletal and a skeletal lineage. At the 18-dpa time point, differentiated dermal and interstitial fibroblasts were not yet identifiable, consistent with live imaging data showing the late differentiation of these lineages (*12*). Notably, both *Col1a2* descendant blastema cells (Fig. 4C, mostly periskeleton derived) and *Prrx1* descendant blastema cells (Fig. 4L, mostly fibroblastic CT derived) had an expression profile resembling that of limb bud progenitors. Taking these results together, we conclude that cells from multiple CT compartments funnel into an undifferentiated progenitor, with the tissue of origin biasing the spatial contribution of the cells.

CT lineages reemerge through multipotent progenitors

To reconstruct the reestablishment of each CT cell type in the upper arm regenerate, we performed high-throughput droplet-based scRNA-seq of labeled *Prrx1:Cre-ER;Caggs:lp-Cherry* descendants in the late blastema at three time points, including 18 dpa (9939 cells), 25 dpa (9019 cells), and 38 dpa (2861 cells) (Fig. 5A and table S9). We inferred differentiation trajectories and pseudotemporal cell relationships (*24*) from diffusion map embedding (*17*) and identified a trifurcated path, where multipotent blastema progenitors expressing embryonic limb and cell cycle markers (e.g., *Prdx2*, *Nrep*, and *Ccnb1*) branch off into a nonskeletal lineage or a skeletal lineage that then bifurcates into either cartilage or bone (Fig. 5, B and C). We observed temporal differences in the lineage progression, as progenitor cells are present only at two earlier time points (18 and 25 dpa), whereas cells on the nonskeletal branch are found at all three time points. A skeletal precursor state is found mainly at 18 dpa, before cartilage and bone start to differentiate at 25 dpa and 38 dpa, respectively (see also fig. S11B). We next analyzed the heterogeneity in the nonskeletal branch and found that the main

nonskeletal lineages (periskeletal cells, tenocytes, and fibroblastic CT cells) present in the adult uninjured tissue had started to reemerge by 38 dpa (Fig. 5D and fig. S11, C and D). These data suggest that, at the transcriptome level, the progenitor pool between 18 and 25 dpa is still relatively homogeneous and that these progenitor cells diversify sometime after 25 dpa into diverse CT lineages.

We further sought to validate our molecular analysis by clonal lineage tracing using a Brainbow transgenic animal, which allowed us to determine the spectrum of cell types formed from single CT cells. To establish clonal tracing conditions, recombination was induced in mature limb cells and examined after 7 days. To identify color combinations showing appropriately low occurrence for clonal analysis, clone pairs were identified as infrequently occurring adjacent sisters derived from a cell division, and their distribution in color and hue-saturation (HS) space was determined (fig. S11E). From these data, a rare recombinant type that mapped in the blue color range was identified (fig. S11F). Examination of nine different source zone equivalents in the mature tissue identified 2, 1, 0, 1, 1, 0, 0, 2, and 0 blue cells per zone (fig. S11G). Amputated limbs were allowed to regenerate for 25 days (Fig. 6A) and then examined for the presence of blue clones. Three out of 10 limbs contained a blue clone, whereas the other seven limbs showed no blue clone, confirming its rare occurrence (Fig. 6, B and D). The frequency distribution in color space of the blue cells in each regenerate was consistent with clonal origin (Fig. 6, C and E; for more examples, see fig. S11, H and I). Assignment of the blue cells from within each limb to CT subtypes revealed contributions of clonal descendants to skeletal, periskeletal, fibroblastic, and tendon cells (Fig. 6B). These lineage tracing data confirm the molecular profiling conclusions that limb blastema cells acquire a limb bud progenitor identity and form a multipotent CT progenitor (*25*, *26*).

Summary

The molecular understanding of blastema formation was previously limited by the inability to identify and isolate blastema precursor cells in the adult tissue. We have demonstrated the importance of genetically marked transgenic axolotl strains for isolating blastema precursor cells from adult limb tissue, and we have molecularly profiled these cells by using single-cell transcriptomic methods. This profiling has indicated that CT cells express adult phenotypes that are lost upon the induction of regeneration and funnel into an embryonic limb bud–like phenotype that includes multipotency within the CT lineage (*25*, *26*). The molecular reprogrammability of adult cells to cells of embryonic limb potential capable of orchestrating complex limb morphogenesis has clear implications for future prospects in regenerative engineering.

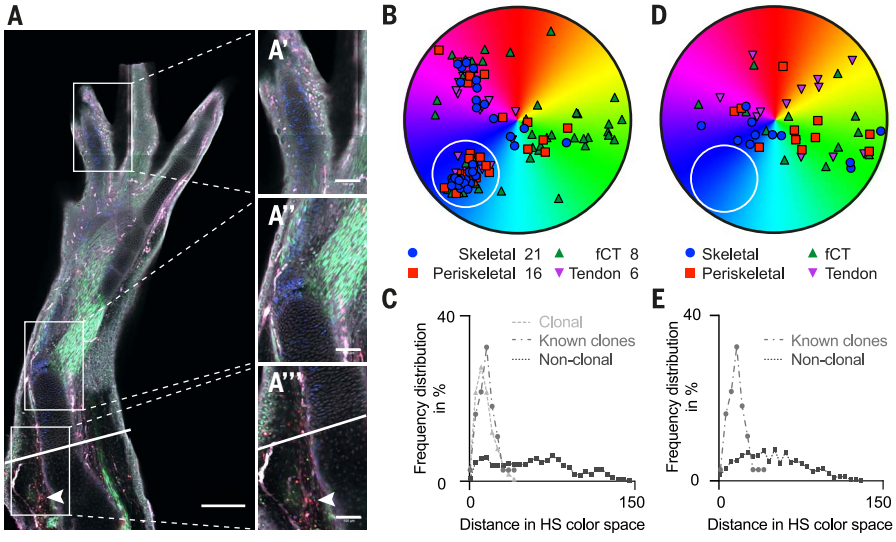


Fig. 6. Brainbow clonal analysis confirms multilineage potential of CT cells upon limb regeneration. (**A**) Image of a regenerated Brainbow axolotl limb. A presumptive clone of blue cells is observed throughout the limb, from the digit tip (A′), the elbow (A″), and the amputation plane (solid line) at the injection site (arrowhead) (A‴). Scale bars, (A) 300 μm; (A′ to A‴), 100 μm. (**B**) HS color distribution of cells from a representative image, including the presumptive clone of blue cells (white circle). Multiple cells of each CT subtype are represented. Similar distributions were observed in 3 of 10 analyzed samples (for examples, see fig. S11H). (**C**) Frequency distribution in HS color space, calculated by using the formula in fig. S11E, for known clonally related cells (fig. S11, F and G) (triangles), presumptive clonally related cells [the blue cells in the white circle in (B)] in a regenerated limb (circles), and non–clonally related cells in a regenerate (squares). The frequency distribution of the presumptive clone is indistinguishable from that of known clones (Kruskal-Wallis analysis and Dunn's multiple comparison). (**D**) Example of HS color distribution of cells from a representative image lacking a discreet cluster of blue cells (white circle). Similar distributions were observed for 7 of 10 analyzed samples (for examples, see fig. S11I). (**E**) Frequency distribution as in (C) for the sample shown in (D). Because of the lack of identification of a clonally related subset, no presumptive clone could be mapped.

Materials and methods

Axolotl husbandry, transgenesis, and 4-OH tamoxifen treatment

Axolotls (*Ambystoma mexicanum*) were bred in MPI-CBG, CRTD, and IMP facilities. All animal handling and surgical procedures were carried out in accordance with local ethics committee guidelines. Animal experiments were performed as approved by the State Authorities Saxony and the Magistrate of Vienna. "White" refers to a non-transgenic d/d strain of axolotl that has white skin due to the absence of melanocytes. Animals were anesthetized in 0.03% benzocaine (Sigma) before amputation and surgery. Reference (*27*) describes axolotl husbandry, transgenesis, and 4-hydroxy tamoxifen (4-OHT) treatment in detail.

For generating transgenic animals, *Prrx1* (*10*), *Col1A2* (*28*), and *Col2A1* (*29*) driver elements were cloned at the 5′ end of the *TFPnls-T2A-ERT2-Cre-ERT2* (*ER-Cre-ER*) or *TFPnls-T2A-Cre-ERT2* (*Cre-ER*) cassette with flanking SceI sites. TFPnls refers to nuclear teal fluorescent protein. The *Col1A2* 1.5-kb promoter and the *Prrx1* 2.4-kb enhancer/promoter from the mouse are kind gifts from George Bou-Gharious and Malcolm Logan, respectively. Constructs used for generating transgenic axolotl driver lines are available at Addgene (111150, 111151, 111152). *Caggs:LoxP-eGFP-3polyA-LoxP-Cherry* (*Caggs:lp-Cherry*) transgenic animals and the *Col2A1* promoter were described in previous publications (*29*, *30*). Transgenic driver animals were generated as described previously by using SceI meganuclease (*27*). General information on the founder animals can be found in table S1. In order to perform tracing, driver lines were crossed with *Caggs:lp-Cherry* reporter animals to obtain double transgenic progeny. We believe that our *Caggs:lp-Cherry* reporter line carries multiple insertion copies of the cassette and thus produces offspring with either one or more than one copy of *Caggs:lp-Cherry*. Occasionally, we obtained animals that showed better conversion efficiency than the rest of the animals. Thus, for each experiment, animals were carefully chosen to ensure that they had a similar level of green fluorescent protein (GFP) intensity before conversion. All tracing and transplantation experiments involving transgenic animals were performed with F_1 or later generations.

4-OHT treatment was done as described previously (*27*) and as presented in table S2. Briefly, small animals were treated with 1 to 2 μM 4-OHT by bathing, and large animals were intraperitoneally injected with 5 μl of 10-mg/ml 4-OHT per gram of body weight of the animals.

Immunohistochemistry and microscopy

Limb cryosections of 10 μM thickness were rehydrated and permeabilized by using phosphate-buffered saline (PBS) with 0.3% Tween-20 and then blocked for 1 hour with PBS with 0.3% Triton-X100 and 2% normal horse serum (Vector Laboratories # S-2000). Slides were then incubated overnight with primary antibodies in blocking buffer in a humidified chamber. The next day, slides were washed with PBS–0.3% Tween-20 and then incubated with secondary antibodies for 2 hours, washed with PBS–0.3% Tween-20, and mounted with mounting medium (Vector Laboratories # H-1000). The cell nuclei were stained with Hoechst 33342 (Sigma; final concentration, 0.5 μg/ml). Mosaic images of sections were acquired on a Zeiss Axio-observer.z1 by using a 20× apochromatic lens with 0.8 numerical aperture (NA) and Axiovision software or Zen2 (Zeiss).

Antibodies

The primary antibodies used in these studies were anti-COL1A2 (DSHB, Sp1.D8), PAX7 (MPI-CBG antibody facility), MHC (monoclonal antibody 4A1025, a kind gift from S. Hughes), β-3TUB (R&D#MAB1195), MBP (Genetex#GTX761141), anti-HOXA11 (MPI-CBG antibody facility), and anti-PRRX1 antibodies (MPI-CBG antibody facility). Secondary antibodies used in these studies were procured from Molecular Probes.

Anti-PRRX1 antibodies

For raising polyclonal anti-PRRX1 antibodies, a DNA fragment of the axolotl *Prrx1* coding for the N-terminal amino acids 1 to 101 was cloned to generate a glutathione *S*-transferase (GST) fusion protein (GST-AxPRRX1-N) and a maltose binding protein (MBP) fusion protein (MBP-AxPRRX1-N). GST-AxPRRX1-N and MBP-AxPRRX1-N fusion proteins were expressed in *Escherichia coli* (BL21-DE3-pLysS) and purified by using glutathione Sepharose 4B (GE) and amylose resin (NEB), respectively. GST-AxPRRX1-N was injected into the rabbit to generate polyclonal antibodies. The serum was affinity purified against MBP-AxPRRX1-N that was immobilized on NHS-activated Sepharose 4 Fast Flow (GE). Antibodies were dialyzed and concentrated by using Amicon Ultra-15 10K MWCO (Millipore) before usage.

To validate anti-PRRX1 antibodies, we first immunostained sections of a developing axolotl limb bud. As expected for a transcription factor, we noticed strong staining in the nuclei of limb mesenchyme and complete absence of staining in the epidermal cells (Fig. 1A). To test whether anti-PRRX1 can stain CT cells of the mature limb, we performed an LPM transplantation from GFP donor embryos (discussed below). Staining of such LPM-transplanted limbs showed complete colocalization among GFP^+ cells and $PRRX1^+$ stained nuclei (fig. S2A). To further validate the specificity of PRRX1 antibodies to CT, we co-stained PRRX1 with markers of muscle fibers (MHC), muscle satellite cells (PAX7), Schwann cells (MBP), and neurons (β-3TUB) in mature limb sections (fig. S2, B to E). All of these markers showed no colocalization with anti-PRRX1 antibodies, suggesting that anti-PRRX1 antibodies specifically label the CT population in the limb bud and the mature limb. Further, as expected, we found an increase in anti-PRRX1 staining in both upper arm and lower arm blastema upon amputation (Fig. 1A). This suggests that PRRX1 is a pan-CT marker for the limb.

Identification of five subtypes of CT

Collagen 1 alpha 2 (COL1A2) is an ECM protein, and in axolotl, it is expressed in dermal fibroblasts and skeleton cells (*28*). Staining limb sections with COL1A2 antibodies showed specific labeling of basal lamina, tendons, and skeletal and periskeletal structures. Thus, on the basis of co-staining of PRRX1 and COL1A2, along with differential interference contrast (DIC) imaging (which delineated the tissue morphology), we identified the following five distinct CT subtypes (fig. S8): (i) dermis/dermal fibroblasts (dF), cells underneath the basal lamina that are $COL1A2^+$ and $PRRX1^{med}$; (ii) interstitial fibroblasts (iF) (fCT I to III), cells scattered across the limb in or around muscle fibers, nerve fibers, and blood vessels, etc., that are $COL1A2^-$ and $PRRX1^{high}$ or $PRRX1^{med}$; (iii) tenocytes (Tn), cells found as thick bundles of CT cells with elongated nuclei in the interstitial space that are $COL1A2^+$ and $PRRX1^{med}$; (iv) periskeletal cells (Ps), cells at the periphery of the skeleton that are $COL1A2^+$ and $PRRX1^{high}$ or $PRRX1^{med}$; and (v) skeletal cells (Sk), cells that reside inside a cartilaginous structure and are $COL1A2^+$ and $PRRX1^{low}$.

Characterization of transgenic animals

Prrx1:Cre-ER animals showed expression of TFPnls in limb buds, with fluorescence progressively declining as the limb developed (fig. S1B). In fully developed limbs, TFPnls is no longer visible under a stereoscope. 4-OHT treatment of *Prrx1:Cre-ER;CAGGs:lp-Cherry* animals at stage 44 (*21*) (just after hatching) resulted in conversion in limb bud mesenchyme (Fig. 1C). In the developed limbs, these converted cells contributed to the entire CT lineage (Fig. 1D and fig. S1).

In *Col2A1:ER-Cre-ER* animals, TFPnls fluorescence was visible in cartilage structures of the primary body axis, mainly the jaw skeleton, but not in the secondary body axis (limb skeleton). However, 4-OHT treatments of *Col2A1:ER-Cre-ER;CAGGs:lp-Cherry* animals at an early hand morphogenesis stage showed conversion not only in the skeleton of the primary axis but also in the limb (fig. S9).

In *Col1A2:ER-Cre-ER* animals, TFPnls fluorescence was visible in skin and cartilage structures (fig. S7A). In a longitudinal section of a *Col1A2:ER-Cre-ER* limb, TFPnls showed complete colocalization with anti-COL1A2 antibodies in dermal fibroblasts, tendons, and periskeletal and skeletal cells but not in the interstitial fibroblasts (fig. S7B). Similarly, conversion of *Col1A2:ER-Cre-ER;CAGGs:lp-Cherry* animals with 4-OHT showed conversion in dermal fibroblasts, tenocytes, periskeletal cells, and skeletal cells but not in the interstitial cells (fig. S7, B and C).

LPM transplantation and skeleton-LPM transplantation

For LPM transplantation, LPM cells from stage 18 *Caggs:eGFP* embryos were transplanted onto white embryos (*4*). Clean LPM transplantation results in CT labeling. Animals with skeleton only were obtained in this process, probably by an incomplete LPM transplant.

Humerus transplantation and interstitial CT transplantation

For humerus transplantation, 5- to 7-cm-long *Prrx1:Cre-ER;CAGGs:lp-Cherry* embryonic converted transgenic animals and white animals were used. Limbs of both transgenic and white animals were amputated at the elbow. All soft tissues including muscle, nerve, blood vessels, and skin were gently pushed back by using forceps to free the humerus from any attachment, and then the humerus was pulled out. Humeri of transgenic and white animals were swapped and were placed into the corresponding cavity. Later, tissues were trimmed to the distal metaphysis of the upper arm to mimic an upper arm amputation and produce a flat amputation surface. Two kinds of humerus-transplanted animals were obtained: (i) *Prrx1:Cre-ER;CAGGs:lp-Cherry* hosts with unlabeled white humeri and (ii) unlabeled white hosts with *Prrx1:Cre-ER;CAGGs:lp-Cherry*–labeled humeri.

Similarly, *Col1A2:ER-Cre-ER;CAGGs:lp-Cherry*–labeled humeri were grafted into the nonlabeled white host to obtain preferential labeling in the skeleton and periskeletal cells. Such limbs were harvested 15 days later for the analysis of HOXA11 staining.

For interstitial CT transplantation, 5-cm-long *Prrx1:Cre-ER;CAGGs:lp-Cherry* limb bud converted transgenic animals were used as donors. From the amputated limbs from donors, the full thickness of skin was removed and discarded. Next, humeri were separated from soft tissues and discarded. Remaining soft tissues, including labeled interstitial CT cells, were placed in 0.8× PBS. In parallel, white hosts underwent amputation near the distal metaphysis, and humeri were trimmed to produce a flat surface. Some muscle tissues were carefully removed and replaced with the soft tissue obtained from the donor to obtain interstitial CT–transplanted animals.

Five-centimeter-long animals were used for all three transplantation procedures. After transplantation, the hosts were kept anesthetized for another 2 hours in a humidified chamber with 0.005% benzocaine to allow healing. Animals were then carefully transferred to fresh water.

Tracing experiments and analysis

Tracing experiments were carried out with 4- to 5-cm-long animals (table S2). For tracing experiments, animals were amputated at the distal humerus near the metaphysis, and humeri were trimmed to produce a flat amputation surface. Animals were imaged on an Olympus SZX16 fluorescent stereoscope every 5 days to monitor regeneration. At 25 dpa, regenerated limbs were

harvested by amputation near the shoulder and fixed in 4% MEMFA overnight at 4°C. The next day, limbs were washed three times with PBS, dehydrated in 30% sucrose, and embedded in Tissue-tek O.C.T. compound (Sakura). Limbs were sectioned and stained as described earlier. Cross-section analysis and identification of each subtype were done on the basis of DIC morphology and immunostaining of PRRX1 and COL1A2 as described earlier. In all cases, limbs were analyzed at four proximal-distal positions: namely, UA-mature (upper arm, mature), UA-callus (upper arm, transition zone), UA-regenerate (upper arm, regenerate), and LA-regenerate (lower arm, regenerate).

For *Col2A1:ER-Cre-ER;CAGGs:lp-Cherry* tracing, converted cells were counted only in the skeletal compartment because no converted cells were observed in any other CT subtype. For humerus-transplanted animals, counting was done in periskeletal and skeletal compartments. For *Col1A2:ER-Cre-ER;CAGGs:lp-Cherry* and *Prrx1:Cre-ER;CAGGs:lp-Cherry* (conversion in the mature limb) tracing, converted cells in all five subcompartments were counted. For *Col1A2:ER-Cre-ER;CAGGs:lp-Cherry* and *Prrx1:Cre-ER;CAGGs:lp-Cherry* (conversion in the mature limb) analysis, cross sections were made on sister slides and stained with COL1A2 and PRRX1 antibodies to identify the presence of total CT cells and converted CT cells in each compartment. For *Prrx1:Cre-ER;CAGGs:lp-Cherry* (conversion in the mature limb) analysis, fold enrichment was calculated as a ratio of percentages of labeled cells in the regenerate over the percentage of soft CT in the mature upper arm limb (preamputation zone). Statistical analyses were performed by using GraphPad Prism 6.0 (GraphPad Software).

Axolotl cell cycle kinetics

The somatic axolotl cell cycle has been measured in the range of 53 to 103 hours (*13*, *14*), which is related to the massive genome (*31*). This means that by 11 days of regeneration, a maximum of 2.5 to 5 cycles have occurred, which translates to a 6- to 32-fold expansion in cell number (see below for actual cell counts). After limb amputation, cells within several millimeters respond to the injury by initiating S phase. Previous bulk molecular profiling comparing wounding with limb amputation across very fine time points determined that the common injury response lasts for 3 days (*15*).

Blastema founding population

Two important pieces of information came from previous live imaging of bulk CT cells during digit tip regeneration with the use of Brainbow transgenics (*12*):

(i) The majority of CT cells (except cartilage) migrate into the blastema, excluding the selection of a rare cell type and validating the FACS approach.

(ii) Cells within 500 μm migrate into the blastema, allowing us to calculate the founding population.

Cell numbers in the axolotl blastema and the founding population

The 11-day blastema has on average 45,000 PRRX1$^+$ cells, whereas a 500-μm slice of the mature limb harbors on average 7500 PRRX1$^+$ cells, excluding cartilage. This implies, on average, a sixfold expansion of cell number during regeneration, which is consistent with an average 103-hour cell cycle. The cell cycle numbers combined with these cell counts imply that an embryonic cell–like precursor would have to be present conservatively at 20 to 100% of the mature cell population.

Brainbow clonal analysis

For clonal tracing, 5-cm-long Brainbow axolotls (*12*) were injected with Tat-Cre protein to indelibly label cell lineages with different fluorescent proteins. Limb amputations were performed 7 days postinjection, just in front of converted cells. Limbs were allowed to regenerate for 25 days and then fixed by using MEMFA. Limbs were cleared by using a 1-propanol$_{pH9}$/Eci clearing approach (*32*). *Z*-stacks were acquired on an inverted Zeiss LSM780 equipped with a 10×/0.3 EC Plan-Neofluar objective (Carl Zeiss Microscopy GmbH, Germany). Zen 2.3 SP1 (Zen black/64 bit) was used for image acquisition and automatic stitching of images. Representative RGB slices were acquired by using FIJI (*33*) and processed as described previously (*34*) with minor changes. Hue and saturation values were extracted by using the 5 by 5 Average option for the Magic Wand tool in Adobe Photoshop CS6. Clonal analysis was performed essentially as described previously (*35*) with modifications. To determine the distance in color space between known clones, control limbs were left unamputated and surveyed for distinct pairs of cells with similar color, location, and morphology. These cells are assumed to be clonally derived from each other and provide a measure for the expected variance in color space for clonally related cells. To determine the expected prevalence of clonal distribution in color space, unamputated control samples were divided into 350-μm regions and analyzed for clonal diversity. These regions represent the 500-μm area that contributes all the cells required for limb regeneration after taking into account the clearing-induced shrinkage of ±30 to 40%. Regenerated limbs were analyzed by selecting both presumptive clones and non–clonally related cells, while recording cell type, hue, and saturation. Kruskal-Wallis analysis and Dunn's multiple comparisons test were used to determine differences in clonal analysis. Frequency distributions were used to graphically display samples. Outliers were removed by using ROUT (Q=1%). Three unamputated control samples and 10 regenerated samples were used for analysis. Statistical analysis was performed using Prism7 (version 7.0c).

Dissociating and preparing axolotl upper arm tissue for scRNA-seq experiments

Six to 10 blastemas, limb buds, or uninjured upper forelimbs were dissociated as a pool for each scRNA-seq experiment. Blastema samples and uninjured tissues were dissociated in 500 μl of 1× Liberase (Roche) diluted in 0.8× PBS$^-$ (PBS without Mg^{2+} or Ca^{2+}) supplemented with 0.5-U/μl deoxyribonuclease I (DNase I) (Thermo Scientific) at room temperature for ~30 min. Mechanical disruption by forceps as well as smooth pipetting was applied during incubation. Then, the cell suspension was filtered through a 30-μm-diameter strainer to generate a single-cell suspension. The filter was washed three times with 500 μl of 0.8× PBS$^-$, and either the obtained cell suspension was immediately subjected to FACS or cells were collected by centrifugation at 300 relative centrifugal force (rcf) for 5 min and again suspended/washed in 1 ml of 0.8× PBS for the 10× Genomics experiments. Centrifugation and washing were repeated once more, and cells were afterward again filtered through a 30-μm-diameter strainer. Cells were collected in 50 μl of supernatant and counted manually. For C1 experiments, cells were sorted by FACS in a single tube and centrifuged at 300 rcf for 5 min, after which they were resuspended in ~15 μl of 0.8× PBS and then kept on ice. Cells were counted manually. Limb bud samples were dissociated in 100 μl of 1× Liberase (Roche) in 0.8× PBS$^-$ with disrupting and smooth pipetting at room temperature for 30 min. Afterward, the cell suspension was diluted with 200 μl of 0.8× PBS$^-$ and filtered two times through a 30-μm-diameter strainer to generate a single-cell suspension that was manually counted and directly used for single-cell experiments. Limb field tissues were dissociated in 0.8× PBS$^-$ while EDTA was added until an easy dissociation with pipette tips could be achieved.

Design of scRNA-seq experiments

As an uninjured reference state and to illuminate general axolotl limb heterogeneity, scRNA-seq experiments using the 10× Genomics platform were performed without any cell type enrichment. In contrast, to obtain deeper insights into CT heterogeneity, scRNA-seq experiments with the uninjured mature upper forelimb were performed by using FACS enrichment. To investigate the process of blastema formation, we collected independent scRNA-seq blastema data along a time course at 3, 5, 8, 11, and 18 dpa of axolotl upper arm forelimbs. In all CT-specific experiments, FACS was used to enrich for mCherry$^+$ CT cells from converted *Prrx1:Cre-ER;CAGGs:lp-Cherry* animals. Around 40,000 cells were sorted for each single-cell experiment into 300 μl of 0.8× PBS, and at least two microfluidic C1 chips (Fluidigm, 96 capture sites) were used per experiment. For stage 40 and stage 44 limb bud experiments, dissociated cells were directly used for single-cell isolation on microfluidic chips without FACS enrichment. Wild-type as well as *Prrx1:Cre-ER;CAGGs:lp-Cherry* animals were used for limb bud experiments, and at least two microfluidic C1 chips (Fluidigm, 96 capture sites) were used for each limb bud stage. For the second set of scRNA-seq experiments where we used converted *Col1A2:ER-Cre-ER;CAGGs:lp-Cherry* animals, individual mCherry$^+$ cells were directly sorted by FACS in single wells of a 96-well plate, and SmartSeq2 (*36*) was performed.

Capturing of single cells and preparation of cDNA

Prrx1 converted animals were used for all 10× Genomics experiments. FACS enrichment of the CT was performed for two of the uninjured CT experiments as well as for the blastema samples (18, 25, and 38 dpa). Two experiments on whole upper arm sections were performed without CT enrichment to analyze the complete axolotl limb heterogeneity. Cells were loaded onto the 10× cartridge at various concentrations, with a target of 2000 to 6000 cells per sample depending on the input cell concentration and the total cell number obtained after tissue dissociation. Cell encapsulation, cDNA generation, and preamplification as well as library preparation were performed by using the Chromium Single Cell 3′ v2 reagent kit according to the instruction manual. The isolation of single CT cells after FACS enrichment and the generation of cDNA for *Prrx1* converted samples (uninjured samples and blastemas at 3 to 18 dpa) and limb bud experiments were performed by using the Fluidigm C1 system. Single cells were captured on a medium-sized (10- to 17-μm cell diameter) or large-sized (17- to 25-μm cell diameter) integrated fluidic circuit RNA-seq chip (Fluidigm). Cells were loaded on the chip at a concentration of 200 to 300 cells/μl and imaged by phase-contrast microscopy to assess the number of cells per capture site. Cell capture, cell lysis, reverse transcription, and cDNA amplification were performed on the chip. Reverse transcription of mRNA was performed by using the SMARTer Ultra Low RNA kit v2 for Fluidigm C1 (TaKaRa Clontech), where ERCC (External RNA Controls Consortium) RNA spike-in mix (Ambion) was added to the lysis reaction and processed in parallel to cellular mRNA. The Advantage2 polymerase chain reaction (PCR) kit (TaKaRa Clontech) was used for cDNA preamplification. For Col1a2 converted samples (uninjured samples and blastema at 11 dpa), single cells were isolated and cDNA was generated by performing SmartSeq2 (*37*). Briefly, individual cells were sorted in 4 μl of lysis mix [0.1% Triton X-100, 2.5 mM (each) deoxynucleoside triphosphate, and 2.5 μM dT30 reverse transcription primer] in single wells of a 96-well plate and snap frozen on dry ice. Generation of cDNA and its amplification exactly followed SmartSeq2 instructions, with application of 19× PCR cycles for blastema cells and 22× for uninjured mature cells. Similarly, stage 28 limb field samples were processed; however, 1 μl containing a single cell was pipetted into 3 μl of lysis mix, and snap freezing was not applied. For preamplifying cDNA, 18× PCR cycles were applied.

RNA-seq library construction and cDNA sequencing

A high-throughput electrophoresis-based fragment analyzer (Fragment Analyzer, Advanced Analytical Technologies) was used to assess the single-cell cDNA fragment size distribution and its concentration of every single cell for C1 and SmartSeq2 experiments. Illumina libraries were constructed by using the Illumina Nextera XT DNA sample preparation kit according to the protocol supplied by Fluidigm. Up to 192 libraries were pooled (3 μl each) and purified with SPRI beads. Library concentration and size distribution were assessed on an Agilent Bioanalyzer and with Qubit double-stranded DNA high-sensitivity assay kits and a Qubit 2.0 fluorometer. Each cell was paired-end sequenced (100 base reads) on an Illumina HiSeq 2500 to a depth of 2 to 5 million reads, and base calling, adaptor trimming, and de-multiplexing were performed as described previously (*38*, *39*). 10× Genomics sample libraries were sequenced on individual lanes on an Illumina HiSeq 2500, and base calling, adaptor trimming, and de-multiplexing of single cells were performed by using 10× Genomics Cell Ranger 2.1 software.

Processing and analysis of scRNA-seq data

Raw reads for C1 and SmartSeq2 experiments were processed by using a custom script and pseudoaligned to the axolotl transcriptome (*40*) by using Kallisto with default settings (*41*). Transcript levels were quantified based on the pseudoalignment as transcripts per million reads (TPM) generated by Kallisto. Single-cell reads obtained with 10× Genomics were aligned to the axolotl transcriptome by using STAR (*42*) implemented in the 10× Genomics Cell Ranger 2.0 software, which generated absolute unique molecular identifier counts. Expression values (TPM/counts) of different isoforms of the same gene were summed afterward by using custom Perl scripts. We excluded cells for which fewer than 1000 genes were detected, and ribosomal protein genes were excluded from all datasets. For limb bud samples, cells were excluded if they did not express *Prrx1* (TPM > 0). Expression levels were converted to the log space by taking the $\log_2$(TPM). R studio (*43*) was used to run custom R scripts for performing principal components analysis (PCA) (FactoMineR package) and hierarchical clustering (stats package) and for constructing heat maps, correlation plots, scatter plots, violin plots, dendrograms, bar graphs, or histograms. Generally, ggplot2 and gplots packages were used to generate data graphs. The Seurat package (*44*) implemented in R was used to identify cell clusters and perform differential gene expression analysis on the basis of tSNE. Diffusion maps (*17*) (destiny/dpt packages) were used to analyze cell lineage relationships. Covariance network analysis and visualizations were done by using igraph implemented in R (*45*). Gene ontology (GO) enrichment analyses were performed by using DAVID Bioinformatics Resources 6.7 (*46*). A list of transcription factors was obtained from the online animal transcription factor database AnimalTFDB (*47*).

REFERENCES AND NOTES

1. J. P. Brockes, A. Kumar, Comparative aspects of animal regeneration. *Annu. Rev. Cell Dev. Biol.* **24**, 525–549 (2008). doi: 10.1146/annurev.cellbio.24.110707.175336; pmid: 18598212
2. E. Nacu, E. M. Tanaka, Limb regeneration: A new development? *Annu. Rev. Cell Dev. Biol.* **27**, 409–440 (2011). doi: 10.1146/annurev-cellbio-092910-154115; pmid: 21801016
3. K. Muneoka, W. F. Fox, S. V. Bryant, Cellular contribution from dermis and cartilage to the regenerating limb blastema in axolotls. *Dev. Biol.* **116**, 256–260 (1986). doi: 10.1016/0012-1606(86)90062-X; pmid: 3732605
4. M. Kragl *et al.*, Cells keep a memory of their tissue origin during axolotl limb regeneration. *Nature* **460**, 60–65 (2009). doi: 10.1038/nature08152; pmid: 19571878
5. D. A. Dunis, M. Namenwirth, The role of grafted skin in the regeneration of X-irradiated axolotl limbs. *Dev. Biol.* **56**, 97–109 (1977). doi: 10.1016/0012-1606(77)90157-9; pmid: 320068
6. B. M. Carlson, The effects of rotation and positional change of stump tissues upon morphogenesis of the regenerating axolotl limb. *Dev. Biol.* **47**, 269–291 (1975). doi: 10.1016/0012-1606(75)90282-1; pmid: 1204936
7. E. Nacu *et al.*, Connective tissue cells, but not muscle cells, are involved in establishing the proximo-distal outcome of limb regeneration in the axolotl. *Development* **140**, 513–518 (2013). doi: 10.1242/dev.081752; pmid: 23293283
8. M. J. Pescitelli Jr., D. L. Stocum, The origin of skeletal structures during intercalary regeneration of larval *Ambystoma* limbs. *Dev. Biol.* **79**, 255–275 (1980). doi: 10.1016/0012-1606(80)90115-3; pmid: 7000577
9. T. Nohno *et al.*, A chicken homeobox gene related to *Drosophila paired* is predominantly expressed in the developing limb. *Dev. Biol.* **158**, 254–264 (1993). doi: 10.1006/dbio.1993.1184; pmid: 8101172
10. M. Logan *et al.*, Expression of Cre recombinase in the developing mouse limb bud driven by a *Prxl* enhancer. *Genesis* **33**, 77–80 (2002). doi: 10.1002/gene.10092; pmid: 12112875
11. A. Satoh, D. M. Gardiner, S. V. Bryant, T. Endo, Nerve-induced ectopic limb blastemas in the axolotl are equivalent to amputation-induced blastemas. *Dev. Biol.* **312**, 231–244 (2007). doi: 10.1016/j.ydbio.2007.09.021; pmid: 17959163
12. J. D. Currie *et al.*, Live imaging of axolotl digit regeneration reveals spatiotemporal choreography of diverse connective tissue progenitor pools. *Dev. Cell* **39**, 411–423 (2016). doi: 10.1016/j.devcel.2016.10.013; pmid: 27840105
13. C. D. Vincent, F. Rost, W. Masselink, L. Brusch, E. M. Tanaka, Cellular dynamics underlying regeneration of appropriate segment number during axolotl tail regeneration. *BMC Dev. Biol.* **15**, 48 (2015). doi: 10.1186/s12861-015-0098-1; pmid: 26647066
14. H. Wallace, M. Maden, The cell cycle during amphibian limb regeneration. *J. Cell Sci.* **20**, 539–547 (1976). pmid: 1270529
15. D. Knapp *et al.*, Comparative transcriptional profiling of the axolotl limb identifies a tripartite regeneration-specific gene program. *PLOS ONE* **8**, e61352 (2013). doi: 10.1371/journal.pone.0061352; pmid: 23658691
16. L. Haghverdi, F. Buettner, F. J. Theis, Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* **31**, 2989–2998 (2015). doi: 10.1093/bioinformatics/btv325; pmid: 26002886
17. L. Haghverdi, M. Büttner, F. A. Wolf, F. Buettner, F. J. Theis, Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845–848 (2016). doi: 10.1038/nmeth.3971; pmid: 27571553
18. G. P. Flowers, L. D. Sanor, C. M. Crews, Lineage tracing of genome-edited alleles reveals high fidelity axolotl limb regeneration. *eLife* **6**, e25726 (2017). doi: 10.7554/eLife.25726; pmid: 28917058
19. J. Kang *et al.*, Modulation of tissue repair by regeneration enhancer elements. *Nature* **532**, 201–206 (2016). doi: 10.1038/nature17644; pmid: 27049946
20. Y. Kawakami *et al.*, Wnt/beta-catenin signaling regulates vertebrate limb regeneration. *Genes Dev.* **20**, 3232–3237 (2006). doi: 10.1101/gad.1475106; pmid: 17114576
21. N. P. Bordzilovskaya, T. A. Dettlaff, S. T. Duhon, G. M. Malacinski, "Developmental-stage series of axolotl embryos," in *Developmental Biology of the Axolotl*, J. B. Armstrong, G. M. Malacinski, Eds. (Oxford Univ. Press, 1989), pp. 201–219.
22. E. Nacu, E. Gromberg, C. R. Oliveira, D. Drechsel, E. M. Tanaka, FGF8 and SHH substitute for anterior-posterior tissue interactions to induce limb regeneration. *Nature* **533**, 407–410 (2016). doi: 10.1038/nature17972; pmid: 27120163
23. A. Makanae, K. Mitogawa, A. Satoh, Implication of two different regeneration systems in limb regeneration. *Regeneration (Oxford)* **1**, 1–9 (2014). doi: 10.1002/reg2.16; pmid: 27499860
24. R. G. Parra, N. Papadopoulos, L. Ahumada-Arranz, J. El Kholtei, B. Treutlein, J. Soeding, Reconstructing complex lineage trees from scRNA-seq data using MERLoT. bioRxiv 261768 [Preprint]. 8 February 2018. doi: 10.1101/261768

25. C. G. Arques, R. Doohan, J. Sharpe, M. Torres, Cell tracing reveals a dorsoventral lineage restriction plane in the mouse limb bud mesenchyme. *Development* **134**, 3713–3722 (2007). doi: 10.1242/dev.02873; pmid: 17715176
26. R. V. Pearse 2nd, P. J. Scherz, J. K. Campbell, C. J. Tabin, A cellular lineage analysis of the chick limb bud. *Dev. Biol.* **310**, 388–400 (2007). doi: 10.1016/j.ydbio.2007.08.002; pmid: 17888899
27. S. Khattak *et al*., Optimized axolotl (*Ambystoma mexicanum*) husbandry, breeding, metamorphosis, transgenesis and tamoxifen-mediated recombination. *Nat. Protoc.* **9**, 529–540 (2014). doi: 10.1038/nprot.2014.040; pmid: 24504478
28. G. Bou-Gharios *et al*., A potent far-upstream enhancer in the mouse pro alpha 2(I) collagen gene regulates expression of reporter genes in transgenic mice. *J. Cell Biol.* **134**, 1333–1344 (1996). doi: 10.1083/jcb.134.5.1333; pmid: 8794872
29. R. Kerney, B. K. Hall, J. Hanken, Regulatory elements of *Xenopus col2a1* drive cartilaginous gene expression in transgenic frogs. *Int. J. Dev. Biol.* **54**, 141–150 (2010). doi: 10.1387/ijdb.092848rk; pmid: 19757383
30. S. Khattak *et al*., Germline transgenic methods for tracking cells and testing gene function during regeneration in the axolotl. *Stem Cell Reports* **2**, 243 (2014). doi: 10.1016/j.stemcr.2014.01.017; pmid: 27863220
31. S. Nowoshilow *et al*., The axolotl genome and the evolution of key tissue formation regulators. *Nature* **554**, 50–55 (2018). doi: 10.1038/nature25458; pmid: 29364872
32. W. Masselink, D. Reumann, P. Murawala, P. Pasierbek, Y. Taniguchi, J. A. Knoblich, E. M. Tanaka, Broad applicability of a streamlined ethyl cinnamate-based clearing procedure. bioRxiv 346247 [Preprint]. 13 June 2018. .doi: 10.1101/346247
33. J. Schindelin *et al*., Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012). doi: 10.1038/nmeth.2019; pmid: 22743772
34. Y. A. Pan *et al*., Zebrabow: Multispectral cell labeling for cell tracing and lineage analysis in zebrafish. *Development* **140**, 2835–2846 (2013). doi: 10.1242/dev.094631; pmid: 23757414
35. D. B. Gurevich *et al*., Asymmetric division of clonal muscle stem cells coordinates muscle regeneration in vivo. *Science* **353**, aad9969 (2016). doi: 10.1126/science.aad9969; pmid: 27198673
36. S. Picelli *et al*., Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098 (2013). doi: 10.1038/nmeth.2639; pmid: 24056875
37. S. Picelli *et al*., Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014). doi: 10.1038/nprot.2014.006; pmid: 24385147
38. G. Renaud, M. Kircher, U. Stenzel, J. Kelso, freelbis: An efficient basecaller with calibrated quality scores for Illumina sequencers. *Bioinformatics* **29**, 1208–1209 (2013). doi: 10.1093/bioinformatics/btt117; pmid: 23471300
39. G. Renaud, V. Slon, A. T. Duggan, J. Kelso, Schmutzi: Estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224 (2015). doi: 10.1186/s13059-015-0776-0; pmid: 26458810
40. Axolotl-omics; https://axolotl-omics.org.
41. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016). doi: 10.1038/nbt.3519; pmid: 27043002
42. A. Dobin *et al*., STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013). doi: 10.1093/bioinformatics/bts635; pmid: 23104886
43. RStudio; https://rstudio.com/.
44. R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, A. Regev, Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015). doi: 10.1038/nbt.3192; pmid: 25867923
45. igraph – The network analysis package; http://igraph.sf.net.
46. D. W. Huang, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009). doi: 10.1038/nprot.2008.211; pmid: 19131956
47. H. M. Zhang *et al*., AnimalTFDB: A comprehensive animal transcription factor database. *Nucleic Acids Res.* **40**, D144–D149 (2012). doi: 10.1093/nar/gkr965; pmid: 22080564
48. C. Weinreb, S. Wolock, A. M. Klein, SPRING: A kinetic interface for visualizing high dimensional single-cell expression data. *Bioinformatics* **34**, 1246–1248 (2018). doi: 10.1093/bioinformatics/btx792; pmid: 29228172

ACKNOWLEDGMENTS

We thank B. Borgonovo and D. Drechsel for technical advice; P. Keller for help with anti-PRRX1 antibodies; M. Blaschek, T. Torrecilla Lobos, H. Andreas, B. Gruhl, S. Mögel, and A. Wagner for excellent animal care and breeding; M. Corriero for the axolotl sketch; T. Engelmaier for help with hematoxylin and eosin staining; Z. He, G. Parra, and J. Kelso for computational support; A. Weihmann and B. Schellbach for sequencing; and members of the Tanaka lab and the Treutlein lab and S. Pääbo for helpful discussions. We thank the Core Unit Durchflusszytometrie (CUDZ) at the College of Veterinary Medicine, University of Leipzig, for providing access to FACS. Animal experiments were performed as approved by the State Authorities Saxony and the Magistrate of Vienna. **Funding:** We acknowledge financial support from the Max Planck Society (B.T.), DFG grant TA 274/3-3 (P.M., E.M.T.), DFG TA 274/11-1 (D.K.), DFG/ANR TA274/13-1 (P.M., E.M.T.), an ERC Advanced Investigator award (P.M., S.H., and E.M.T.), an ERC starting grant (B.T.), and FZ111 and institutional support from the IMP (E.M.T.). **Author contributions:** T.G., P.M., D.K., J.G.C., E.M.T., and B.T. designed the study and wrote the manuscript. P.M. generated transgenic axolotl lines with assistance from M.S. and S.K. P.M. performed lineage tracing and immunohistochemistry with help from S.H. P.M. and D.K. performed grafting experiments. P.M. and W.M. performed Brainbow clonal tracing. T.G. performed single-cell experiments with assistance from J.G.C., D.K., and P.M. M.G.-S. performed FACS. T.G. analyzed scRNA-seq data with help from B.T., J.G.C., J.K., P.M., D.K., and E.M.T. S.N. provided assistance with the axolotl transcriptome. J.D.C. generated the HOXA11 antibody. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Single-cell count matrices and fastq files are deposited in NCBI's Gene Expression Omnibus under the accession number GSE106269.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6413/eaaq0681/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S11
Tables S1 to S10
Reference (*49*)

27 September 2017; accepted 5 September 2018
Published online 27 September 2018
10.1126/science.aaq0681

RESEARCH ARTICLE SUMMARY

STRUCTURAL BIOLOGY

Evolutionary shift toward protein-based architecture in trypanosomal mitochondrial ribosomes

David J. F. Ramrath*, Moritz Niemann*, Marc Leibundgut, Philipp Bieri, Céline Prange, Elke K. Horn, Alexander Leitner, Daniel Boehringer, André Schneider, Nenad Ban†

INTRODUCTION: Ribosomes are universally conserved assemblies composed of ribosomal RNA (rRNA) and proteins, where rRNA plays key structural and functional roles. Despite their high degree of conservation, considerable variability is observed in mitochondrial ribosomes (mitoribosomes), with an extreme example found in *Trypanosoma brucei*, the parasite that causes sleeping sickness. In these mitoribosomes featuring the smallest known rRNAs, the severe rRNA reduction is accompanied by the recruitment of many additional proteins.

RATIONALE: The extreme differences in rRNA size and the substitution of many proteins present in all other ribosomes render trypanosomal mitoribosomes an excellent system to reveal the minimal set of rRNA and protein elements essential for ribosomal function and to investigate how ribosomal proteins compensated for the missing rRNA. To address these questions, we determined the atomic structure of the mitoribosome from *T. brucei* using cryo–electron microscopy.

RESULTS: The trypanosomal mitoribosome, composed of 127 ribosomal proteins and two rRNAs, is larger and architecturally more complex than any other ribosome described so far with a molecular weight of 4.5 MDa and a RNA/protein ratio of 1:6. The structural changes are most prominent in the small subunit that exceeds the size of the "large" subunit. Notably, the reduced rRNA is found at the core of the assembly and is encased in a large shell of mitoribosomal proteins. The universally conserved regions are reduced to a minimum and include only a few rRNA and protein elements in the vicinity of key functional regions of the ribosome, namely, the decoding center, the active site, and the binding region for translation factors.

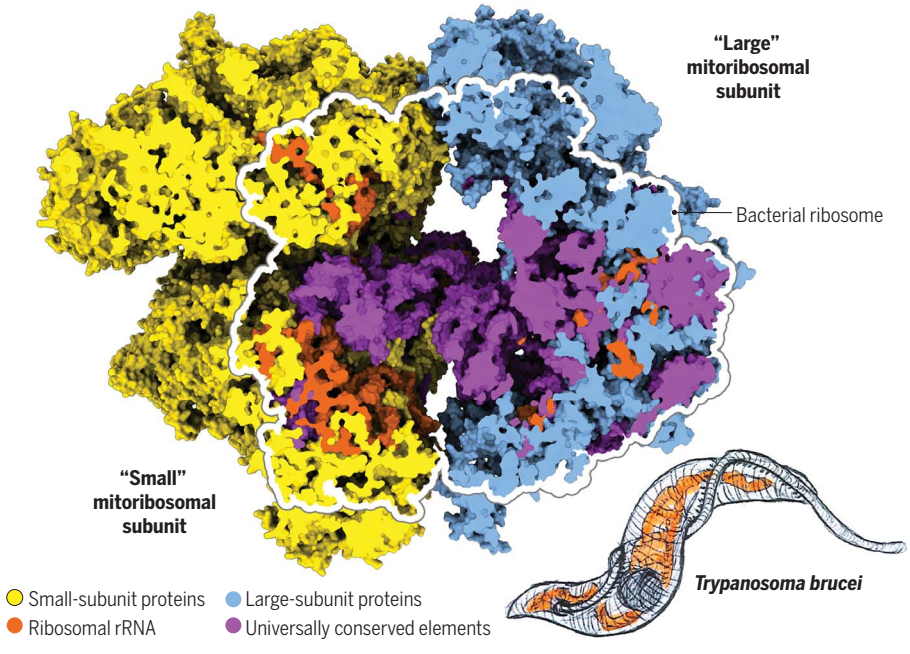
ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aau7735

In contrast to other ribosomes, where the rRNA fold is dominated by a base-paired structure, the proteins take over the architectural role by providing a scaffold for binding of predominantly single-stranded rRNA. The switch to the protein-based architecture is accompanied by a marked increase in the size of conserved ribosomal proteins and the recruitment of novel proteins, including proteins with multiple domains or proteins with homology to various enzymes. Because many proteins contain helical repeat motifs, the assembly contains a disproportionately large fraction of α-helical elements that structurally substitute the reduced rRNA.

Our results show that, because of the extensive remodeling, the trypanosomal translational machinery adopted unusual solutions to accomplish some basic protein synthesis mechanisms. Their nascent polypeptide exit tunnel branches into two exits providing for an interesting possibility that nascent proteins with different characteristics may take different paths. Furthermore, in a subpopulation of isolated small-subunit particles, we observed mitochondrial initiation factor 3 interacting with the decoding center via its unique C-terminal extension, which might compensate for the essential function of initiation factor 1 that is absent in all mitochondria.

CONCLUSION: The unusual architecture and composition of the trypanosomal mitoribosome that we have revealed shows how proteins have taken over a key architectural role from rRNA by forming an autonomous outer shell that serves as a mold for binding flexible single-stranded rRNAs. Moreover, our results show the universally conserved features of ribosomes that are responsible for the most basic functions. Last, structural information on these ribosomes may be helpful for developing new drugs to treat sleeping sickness and other diseases caused by trypanosomes and its relatives. ■



The universally conserved core of the ribosome. In the highly remodeled trypanosomal mitoribosome, where the "small" subunit is larger than the large subunit, proteins took over the key architectural role from the rRNA. The massive protein shell interacts with the extremely reduced rRNA to position functionally critical rRNA elements. The structure helps us define the "minimal" set of conserved rRNA regions and protein components shared by all ribosomes.

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: ban@mol.biol.ethz.ch
Cite this article as D. J. F. Ramrath *et al.*, *Science* **362**, eaau7735 (2018). DOI: 10.1126/science.aau7735

RESEARCH ARTICLE

STRUCTURAL BIOLOGY

Evolutionary shift toward protein-based architecture in trypanosomal mitochondrial ribosomes

David J. F. Ramrath[1]*, Moritz Niemann[2]*, Marc Leibundgut[1], Philipp Bieri[1], Céline Prange[1], Elke K. Horn[2], Alexander Leitner[3], Daniel Boehringer[1], André Schneider[2], Nenad Ban[1]†

Ribosomal RNA (rRNA) plays key functional and architectural roles in ribosomes. Using electron microscopy, we determined the atomic structure of a highly divergent ribosome found in mitochondria of *Trypanosoma brucei*, a unicellular parasite that causes sleeping sickness in humans. The trypanosomal mitoribosome features the smallest rRNAs and contains more proteins than all known ribosomes. The structure shows how the proteins have taken over the role of architectural scaffold from the rRNA: They form an autonomous outer shell that surrounds the entire particle and stabilizes and positions the functionally important regions of the rRNA. Our results also reveal the "minimal" set of conserved rRNA and protein components shared by all ribosomes that help us define the most essential functional elements.

Ribosomal RNA (rRNA) plays key functional and architectural roles in ribosomes. Using electron microscopy, we determined the atomic structure of a highly divergent ribosome found in mitochondria of *Trypanosoma brucei*, a unicellular parasite that causes sleeping sickness in humans. The trypanosomal mitoribosome features the smallest rRNAs and contains more proteins than all known ribosomes. The structure shows how the proteins have taken over the role of architectural scaffold from the rRNA: They form an autonomous outer shell that surrounds the entire particle and stabilizes and positions the functionally important regions of the rRNA. Our results also reveal the "minimal" set of conserved rRNA and protein components shared by all ribosomes that help us define the most essential functional elements.

Ribosomes are two-subunit ribonucleoprotein assemblies responsible for protein synthesis in all organisms and are considered among the most conserved cellular complexes (*1*). Consequently, the sequence of the bacterial small-subunit 16*S* rRNA has traditionally been used to establish evolutionary relationships between organisms (*2*, *3*). Structural information on ribosomes from different kingdoms of life revealed a conserved ribosomal core where the peptidyl transferase center (PTC) and the decoding center are formed by rRNA, whereas surface features vary in terms of protein composition and lengths of rRNA expansion segments to account for different regulatory mechanisms that control translation in different organisms (*4*). So far, the most pronounced structural deviation from bacteria is found in mammalian mitochondria. These mitoribosomes are specialized for the synthesis of membrane proteins and feature a reduced rRNA and increased protein content (*5*, *6*). Notably, an even more extreme rRNA reduction is found in mitochondria of kinetoplastids such as *T. brucei*, *Trypanosoma cruzi*, and *Leishmania* sp., the parasites that, in humans, cause sleeping sickness, Chagas disease, and Leishmaniasis. Their mitoribosomal small-subunit 9*S* and large-subunit 12*S* rRNA (comprising 620 and 1176 nucleotides, respectively, in *T. brucei*) are the smallest known rRNA homologs, reduced to ~40% of the length of canonical eubacterial rRNA (*7*–*9*). The trypanosomal mitochondrial rRNA sequences have an ~80% adenine and uracil content, which hampers the prediction of secondary structure elements that are otherwise present in all rRNAs (*9*). Proteomic studies of affinity-purified trypanosomal mitoribosomes indicate that the rRNA reduction was accompanied by the recruitment of many additional mitoribosomal proteins; however, it was not possible to obtain clear information regarding their size or regarding the number of ribosomal proteins that compose their subunits (*10*–*12*). Nevertheless, these results indicated that trypanosomal mitoribosomes have a unique composition and architecture distinct from any other cytoplasmic and mitochondrial ribosomes (*5*, *6*, *13*, *14*).

The extent of anticipated architectural differences in trypanosomal mitoribosomes render it an excellent system to establish the most conserved set of rRNA and protein elements essential for ribosomal function and to reveal possible changes in the rRNA, which dominates the architecture of all currently investigated ribosomes. Here, we present the cryo–electron microscopy (cryo-EM) structures of both mitoribosomal subunits of the parasitic protozoan *T. brucei* at 3.1- to 3.4-Å resolution in the context of the entire trypanosomal mitoribosome resolved at 7.8 Å. Our reconstructions reveal the unique composition and the unusual protein-based architecture of trypanosomal mitoribosomes.

Results

Sample preparation, structure determination, and model building

Because of the low abundance of kinetoplastid mitoribosomes and their unusual heterogeneity (*10*), they were purified using genetically engineered PTP affinity tags (*15*). However, without knowing the exact composition and architecture of these mitoribosomes, we had to iteratively improve the placement of the tags based on preliminary structural information. From purified mitoribosomes that carried either a large- or small-subunit tag exposed on the solvent accessible side of the particle (fig. S1), we acquired two cryo-EM datasets, which were subsequently processed by several rounds of two-dimensional (2D) and 3D classification steps (fig. S2).

This approach yielded maps of the entire *T. brucei* mitoribosome, its small (Tb-mt-SSU) and large (Tb-mt-LSU) subunits and a distinct class of small subunits in complex with the C-terminal domain (CTD) of initiation factor IF-3 (Tb-mt-SSU•IF-3). For the final high-resolution refinements, individual masks for the Tb-mt-LSU and the Tb-mt-SSU head and body further improved the map quality, resulting in cryo-EM maps at overall resolutions of 3.1 Å for the Tb-mt-SSU head, 3.3 Å for Tb-mt-SSU•IF-3 body, and 3.4 Å for the Tb-mt-LSU (Fig. 1A, figs. S2 and S3, and table S1). Most novel proteins were identified by predicting their sequence based on the side-chain features in EM maps, followed by sequence homology search, 3D modeling, and manual rebuilding. The resulting hits were further corroborated by comparison with the mass spectrometry data that we have collected (data S1). Most of the rRNA segments were readily visible; however, some regions of the 9*S* and 12*S* rRNA at the subunit interface were flexibly disposed, which precluded a molecular interpretation. In the case of the small ribosomal subunit, the complete 9*S* rRNA structure could be determined using the Tb-mt-SSU•IF-3 map, where these areas were well ordered. The missing regions of the large-subunit 12*S* rRNA were modeled for display purposes using the structure of the mammalian mitoribosomal 16*S* rRNA as a guide based on sequence and secondary structure homology (Fig. 1B) (*5*, *6*). On the surface of the mitoribosomal subunits, and at

[1]Department of Biology, Institute of Molecular Biology and Biophysics, Otto-Stern-Weg 5, ETH Zurich, CH-8093 Zurich, Switzerland. [2]Department of Chemistry and Biochemistry, University of Bern, CH-3012 Bern, Switzerland. [3]Department of Biology, Institute of Molecular Systems Biology, Auguste-Piccard-Hof 1, ETH Zurich, CH-8093 Zurich, Switzerland.
*These authors contributed equally to this work.
†Corresponding author. Email: ban@mol.biol.ethz.ch

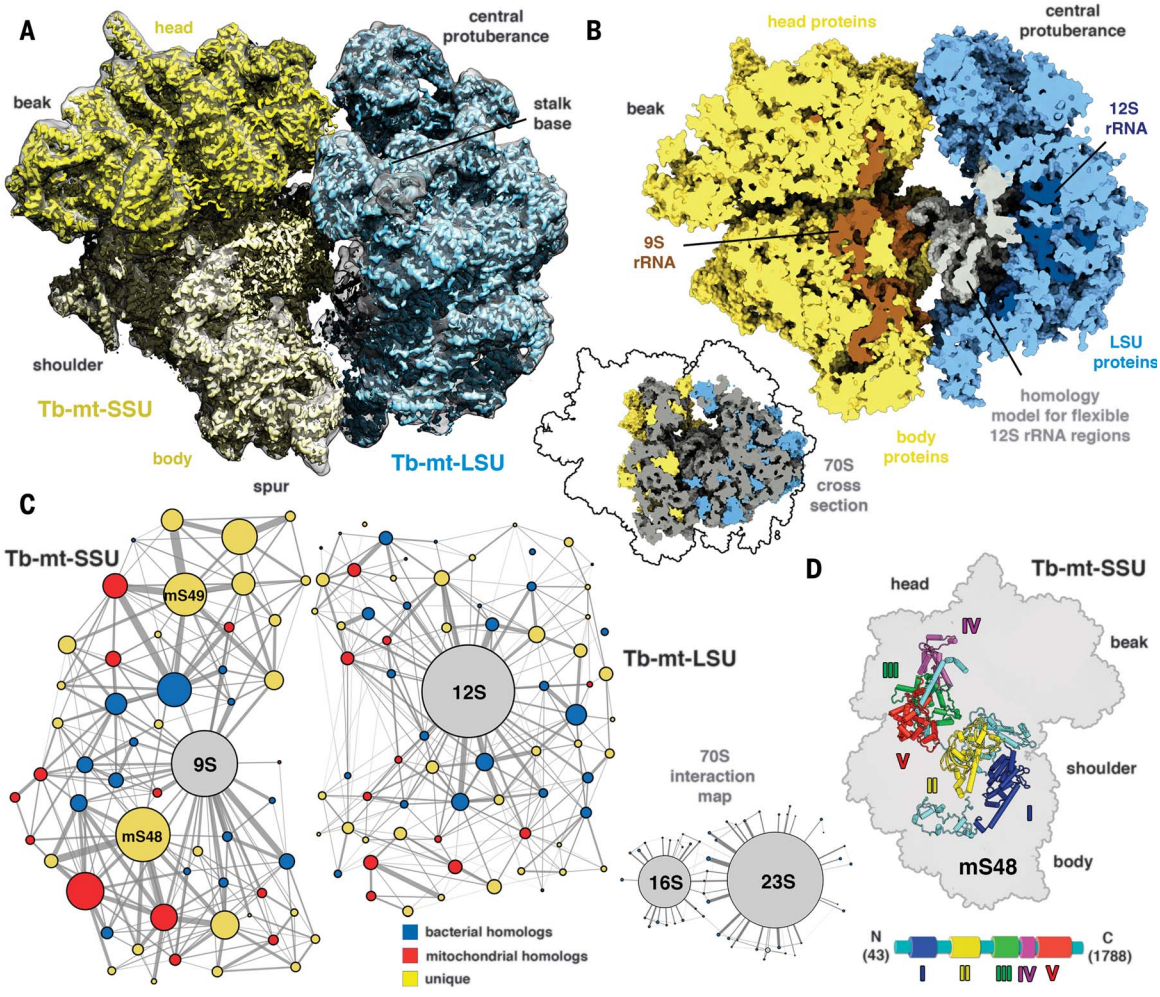


Fig. 1. Overview of the *T. brucei* mitoribosome. (**A**) Cryo-EM reconstruction of the complete trypanosomal mitoribosome shown in transparent gray together with the high-resolution maps of the large subunit (Tb-mt-LSU) in blue and the head and body of the small subunit (Tb-mt-SSU) in dark and bright yellow, respectively. (**B**) Cross section through the mitoribosome with the ribosomal proteins colored as in (A). The 12*S* rRNA homology model is shown in gray, and the atomic models of the 9*S* and the 12*S* rRNA are in brown and dark blue, respectively. Inset: A cross section through the bacterial 70*S* ribosome with the outline of the trypanosomal mitoribosome is shown for comparison. (**C**) Interaction network of all individual elements within the mitoribosome using the indicated color scheme. The size of nodes and thickness of edges correlate with the size of the surface of the individual elements and their interactions. A bacterial 70*S* interaction network (not to scale) is also shown. (**D**) Cartoon representation of mS48 with its individual domains.

the subunit interface, there are several less well-ordered disconnected densities that could not be assigned (fig. S4). The refined subunit structures were docked into the 7.8-Å map of the entire trypanosomal mitoribosome (Figs. 1 and 2 and figs. S2 and S3).

The overall structure of the trypanosomal mitoribosome

The trypanosomal mitoribosome has a diameter of ~385 Å, which makes this ribosome considerably larger and different in appearance than any other ribosomal complex described so far (Fig. 1) (*4*, *14*, *16*). Our structure disagrees with the cryo-EM map previously obtained for a related mitoribosome from *Leishmania tarentolae* (supplementary text) (*17*).

Both trypanosomal mitoribosomal subunits contain pronounced structural features that have not been observed in any other ribosomes: The small subunit harbors a massive beak region that increases the head size to 230 Å, making it almost as large as the body of the small subunit (Fig. 1, A and B). Notably, the small subunit with a calculated molecular weight of ~2.5 MDa is about three times larger than the canonical bacterial small subunit, thereby exceeding even the size of the "large" subunit of the trypanosomal mitoribosome. Also, the overall appearance of the large mitoribosomal subunit is unusual with its massive central protuberance (CP) and the L1 stalk formed entirely by ribosomal proteins.

Both 9*S* and 12*S* rRNAs are found at the core of the trypanosomal mitoribosome encased in a large and highly interconnected shell of mitoribosomal proteins (Fig. 1, B and C). A remarkable feature of the ribosomal protein shell is the size and domain organization of individual proteins such as mS48 or mS49. With a length of 1788 amino acids, mS48 folds into a multidomain protein, in which individual domains are connected by linker regions. This protein contacts numerous ribosomal proteins and extends from the body of the small subunit to the head while contacting the 9*S* rRNA only tangentially (Fig. 1, C and D). This suggests a central organizational role of mS48 in the formation of the protein network of the small subunit, which is in line with the growth defects observed when the protein is ablated by RNA interference (*18*). The second largest protein, mS49, is located in the head of the small subunit and establishes multiple protein-protein interactions that together bury a surface area of ~34,000 Å^2, an area comparable to the one between the bacterial small-subunit 16*S* rRNA and all 30*S* proteins (Fig. 1C and fig. S5) (*19*).

Apart from the 12*S* and 9*S* rRNA, we identified 127 mitoribosomal proteins, 70 in the large and 57 in the small subunit (Fig. 2 and table S2). Roughly half of the proteins are specific for trypanosomes and related species and, according to their molecular weight, were assigned as novel mitochondrial proteins mL67 to mL100 for the Tb-mt-LSU and mS48 to mS74 for the Tb-mt-SSU (table S2). Even among the conserved ribosomal proteins such as uS5m, uS9m, or bL17m, the homology is typically limited to the core domains, whereas the long extensions are specific for trypanosomes and together with the novel proteins amount to ~70% of the total protein mass of the trypanosomal mitoribosome

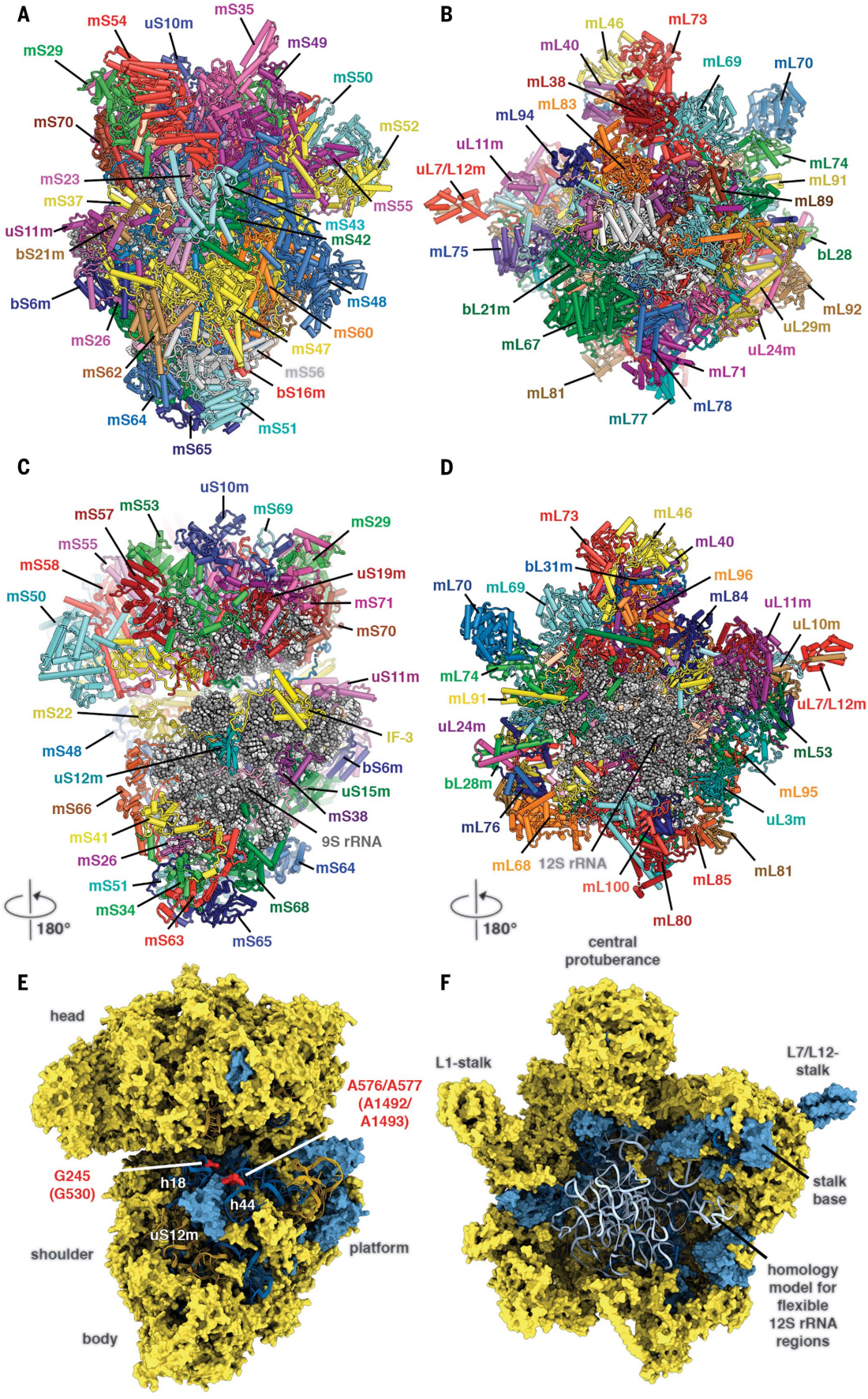


Fig. 2. Atomic models of the small and large mitoribosomal subunits and their universally conserved elements. (**A** to **D**) The individually colored mitoribosomal proteins are shown in cartoon and the rRNAs as gray spheres. The Tb-mt-SSU•IF3 complex (A and C) and the Tb-mt-LSU (B and D) are visualized from the solvent exposed side (A and B) and from the subunit interface (C and D). (**E** and **F**) View on the subunits as in (C) and (D) showing proteins as the surface and rRNAs as ribbons. Universally conserved elements are colored in blue, and the rest is in yellow. The decoding bases of the trypanosomal 9S rRNA (corresponding *E. coli* numbering in parentheses) are indicated in red.

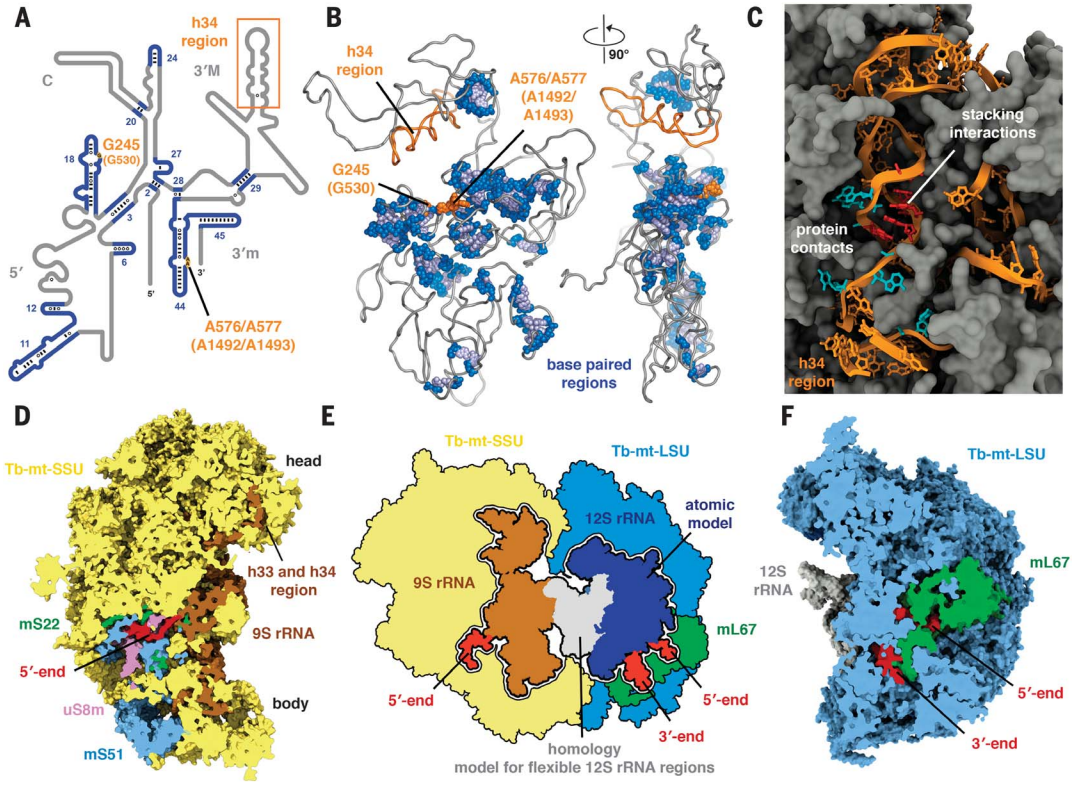


Fig. 3. Features of the 9S and 12S rRNA. (**A** and **B**) Secondary (A) and tertiary (B) structure of the 9S rRNA (gray) illustrating base pairing within residual helical regions (blue) together with the universal conserved decoding bases (orange). (**C**) Close-up view of the 9S rRNA helix h34 (orange cartoon) embedded into the Tb-mt-SSU head proteins (gray surface). (**D** to **F**) Cross sections of the Tb-mt-SSU (D) and Tb-mt-LSU (F) showing the anchoring of the 9S and 12S rRNA termini in the protein shell. The situation in the complete mitoribosome is visualized schematically in (E).

(figs. S6 and S7). Furthermore, trypanosomal mitoribosomal proteins are substantially larger than the bacterial homologs with a ~2.4 times increased median molecular weight. Consequently, the entire mitoribosome has a molecular weight of 4.5 MDa with an RNA/protein ratio of 1:6, which notably differs from the 2:1 ratio found in bacteria (*18*) and the 1:2 ratio found in mammalian mitochondria (*5*, *6*).

The structure of the *T. brucei* mitoribosome in which only a very small number of proteins and rRNA elements are preserved can help us to establish the minimal set of irreplaceable components of this machinery. For the large subunit, this includes the PTC with uL16m next to it and the exit tunnel region formed by uL4m, uL22m, and uL24m (Fig. 2F). Also, the L7/L12 stalk and its base are conserved, formed by the 12*S* rRNA helices H43 and H44 and the protein uL10m (Fig. 2F and fig. S8). The conserved core of the small subunit is even more reduced to the platform region formed by proteins bS6m, uS11m, uS15m, and bS21m (Fig. 2E), as well as the decoding center, where rRNA helices h18 and an extremely shortened h44 are located. These regions of the 9*S* rRNA contain the strictly conserved decoding residues G245 (G530 in *Escherichia coli*) and A576/A577 (A1492/A1493 in *E. coli*), with the sequence corresponding to the G530 loop being the longest conserved nucleotide pattern in the entire trypanosomal mitoribosome (Fig. 2E and fig. S9). The putative transfer RNA (tRNA) binding sites on the small mitoribosomal subunit are surrounded by a number of rRNA nucleotides that are not conserved at a sequence level but occupy equivalent positions in the structure to mediate tRNA interactions (fig. S10). The decoding center rRNA helices are held in place by protein uS12m in a similar manner as in other ribosomes. Trypanosomal uS12m may further play a role in the regulation of ribosome assembly considering that it is, unlike any other mitoribosomal protein, encoded as a cryptogene on the mitochondrial genome (*20*). Thus, its primary transcript is subject to extensive RNA editing, which is required to convert it into a translatable mRNA (fig. S11 and supplementary text) (*21*).

The rRNA of the trypanosomal mitoribosome

Mitochondrial rRNAs of trypanosomes represent markedly reduced versions of the typical rRNA as evidenced by comparing their secondary structure diagrams (figs. S8 and S9). The reduction is also apparent in the functionally important rRNA helices such as H38 in the large subunit and h18 or h44 in the small subunit (Fig. 3 and figs. S8 and S9). Compared to the rRNAs of the bacterial or mammalian mitoribosomal large subunits, many of the stem loop regions of the trypanosomal 12*S* rRNA have been shortened or removed (fig. S8). The changes in the 9*S* small-subunit rRNA are even more pronounced because not only were stem loops reduced but the secondary structure was also extensively remodeled (Fig. 3A and fig S9). In contrast to canonical rRNAs that adopt a compact fold dominated by base-paired double helical secondary structure elements, only ~20% of the 9*S* rRNA sequence is base paired (Fig. 3B). Accordingly, only 10 of 45 helices of the bacterial 16*S* rRNA are conserved in the 9*S* rRNA. In the head of the small subunit, almost all canonical rRNA helices of the 3′ major domain (h30 to h43) are absent or replaced by single-stranded segments of the 9*S* rRNA that are stabilized by either stacking interactions or through contacts with ribosomal proteins. For example, 9*S* rRNA residues 441 to 483 within the h34 region are twisted over a distance of 60 Å following an approximate helical path in the absence of any base pairs (Fig. 3C). Nevertheless, the helix h34 region still occupies the same location on the small subunit, possibly to mediate interactions with mitochondrial elongation factor EF-G1, whose bacterial homolog contacts the minor groove of h34 during tRNA translocation (*22*, *23*).

The interactions between ribosomal proteins and rRNA involve, in some cases, a very unusual topology. For example, the proteins mS53 and

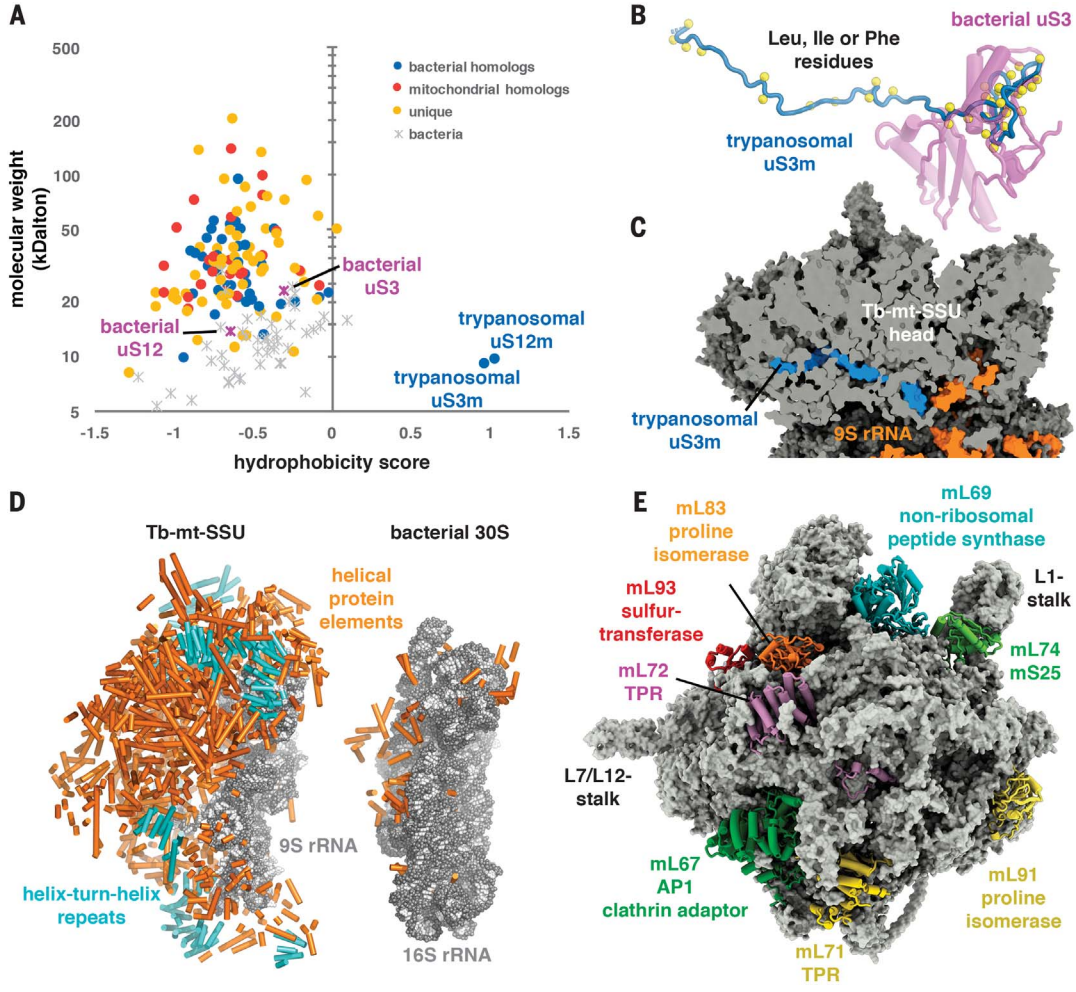


Fig. 4. Proteins of the trypanosomal mitoribosome. (**A**) The molecular weight (log scale) and hydrophobicity score (*31*) of all trypanosomal mitochondrial and bacterial ribosomal proteins (colors as indicated). (**B**) Cartoon representation of trypanosomal protein uS3m (blue) superimposed onto bacterial uS3 (violet). The hydrophobic residues of uS3m are shown as yellow spheres. (**C**) Surface representation of uS3m as an integral part of the Tb-mt-SSU head (proteins in gray and 9*S* rRNA in orange). (**D**) Side views onto the Tb-mt-SSU (left) and bacterial 30*S* small subunit (right) showing α helices of the ribosomal proteins (orange cartoons) and the rRNAs (gray spheres). Helical protein repeats are highlighted in cyan. (**E**) Numerous proteins (cartoon) that share structural homology to non-ribosomal proteins or enzymes are found in the Tb-mt-LSU (gray surface).

mS57 hook the rRNA to the head of the small subunit with their long N-terminal segments threaded through an rRNA loop in the h33 region (fig. S12), suggesting that folding of the 9*S* rRNA is coordinated with the folding and association of ribosomal proteins. This is also evidenced by the fact that the complementary sequences within many 9*S* rRNA regions, which are observed in a single-stranded conformation, would in theory allow the formation of helical structures (such as the equivalents to bacterial helices h7, h15, h22, or h30; fig. S9) (*7*–*9*). Therefore, it is likely that the binding of the 9*S* rRNA to the protein shell keeps it from adopting a secondary structure that would correspond to the lowest energy state predicted for this RNA based on sequence (*7*–*9*).

We observe that the rRNAs of both subunits are anchored in the protein shell (Fig. 3, D to F). This suggests that the massive protein shell represents a cradle that serves as a platform for the folding of the rRNA. The extended single-stranded 5′ end of the 9*S* rRNA is attached to mS51 and trypanosomal-specific elements of uS8m, uS22m, and mS47 (Fig. 3, D and E). In the case of the 12*S* rRNA, both the 5′ and 3′ ends are bound to trypanosomal-specific mL67, the largest Tb-mt-LSU protein (Fig. 3, E and F). These interactions effectively bring together the two ends of the 12*S* rRNA, which emulates the interactions observed in bacteria where the two ends of the 23*S* rRNA are circularized through base pairing of complementary sequences to generate helix H1, which is then processed during maturation (*24*).

Proteins of the trypanosomal mitoribosome

Within the trypanosomal mitoribosome, several functionally important regions of rRNA that play a role in interactions with the mRNA and tRNA substrates or undergo conformational changes during translation have been replaced by protein features. For example, the L1 stalk is solely formed by proteins bL9m, mL70, mL74, and mL91 that protrude away from the large subunit, thereby mimicking the features of the rRNA-based L1 stalk of other ribosomes (Fig. 2, B and D) (*18*). The canonical L1 stalk changes its position by up to 40 Å upon binding and release of tRNAs at the ribosomal E site (*25*). The preservation of structural features of this ribosomal element, although formed exclusively by ribosomal proteins instead of rRNA, suggests an indispensable functional role in interactions with the E-site tRNA and in modulating accompanying conformational changes (*26*).

The trypanosomal mitoribosomal CP is about three times the size of its bacterial counterpart (*18*). It is composed of bL31m, mL38, mL40, mL46, mL73, and mL96, thus representing the first ribosomal CP that is exclusively shaped by protein elements in the absence of any RNA element such as the 5*S* rRNA in bacteria (*18*), the CP-tRNA in the mammalian mitoribosome (*27*, *28*), or an extension of the large-subunit rRNA in yeast mitochondria (Fig. 2, B and D) (*29*). In terms of the protein composition, the trypanosomal CP is dominated by mitochondrial homologs, which suggests that this architectural landmark was already present in ancestral mitoribosomes before the eukaryotes branched into different groups (Fig. 2F and fig. S7).

With the switch to a protein-based architecture, we observe marked changes in the chemical properties of even the most conserved ribosomal proteins, with uS3m and uS12m being striking examples (Fig. 4A). Trypanosomal uS3m is surrounded exclusively by proteins instead of rRNA and shares structural homology with the CTD of bacterial uS3 (Fig. 4B). We identified

uS3m on the mitochondrial genome (table S2 and supplementary text), which is similar to yeast, where uS3m represents the only ribosomal protein encoded by the mitochondrial DNA (*30*). In contrast to typical ribosomal proteins that are rich in positively charged amino acids interacting with rRNA, trypanosomal uS3m is rich in hydrophobic amino acids (23% Ile, 13% Leu, and 13% Phe), resulting in an average hydrophobicity score comparable to membrane proteins (Fig. 4A) (*31*). Therefore, this protein appears to serve as a hydrophobic core in the protein shell (Fig. 4C). The extreme hydrophobic character of uS3m and uS12m may necessitate their cotranslational incorporation into the trypanosomal mitoribosome and may explain why both are encoded on the mitochondrial genome. However, this is unlikely to be the only source of evolutionary pressure to keep these genes on the mitochondrial genome because yeast uS3m has a hydrophilic character similar to bacterial protein (score of −0.786).

The structure of the trypanosomal mitoribosome provides insights into the large class of helical repeat proteins containing penta/tetratricopeptide repeat (P/TPR), HEAT, or ARM motifs (Fig. 4D) (*32*, *33*). The group of P/TPR-containing proteins are characterized as single-stranded RNA binding proteins, frequently found in the chloroplasts and mitochondria. P/TPR proteins were extensively characterized by sequence analysis and biochemical experiments (*33*) and suggested to be part of the trypanosomal mitoribosome (*34*, *35*). Our results reveal the structures and the network of interactions of six P/TPR-containing mitoribosomal proteins (table S2). All of these proteins mediate protein-protein interactions, whereas only some have peripheral contact to rRNA. Presumably, these P/TPR-containing proteins had originally been recruited to single-stranded rRNA segments and maintained their position as integral architectural elements, whereas the surrounding rRNA was reduced (fig. S13 and supplementary text). In general, we observe that the trypanosomal-specific proteins and protein extensions harbor a disproportionally large fraction of α-helical secondary structure elements, many arising from superhelical structures of repetitive helix-turn-helix motifs that architecturally substitute for the reduced rRNA. For the small subunit, the ratio of α helix versus β sheet content is 7.5:1 in trypanosomal mitoribosomes, whereas it is 2:1 in bacterial ribosomes (Fig. 4D).

Many solvent-exposed proteins of the trypanosomal mitoribosome are structurally homologous to nonribosomal proteins or enzymes. The folds of the AP1 clathrin adaptor (mL67), proline isomerase (mL91), sulfur transferase (mL93), and nonribosomal peptide synthase (mL69) are found in the large subunit and those of thiosulfate sulfurtransferase (mS67) and A-kinase anchoring protein (mS64) in the small subunit (Fig. 4E and table S2). However, these proteins with homology to enzymes contain

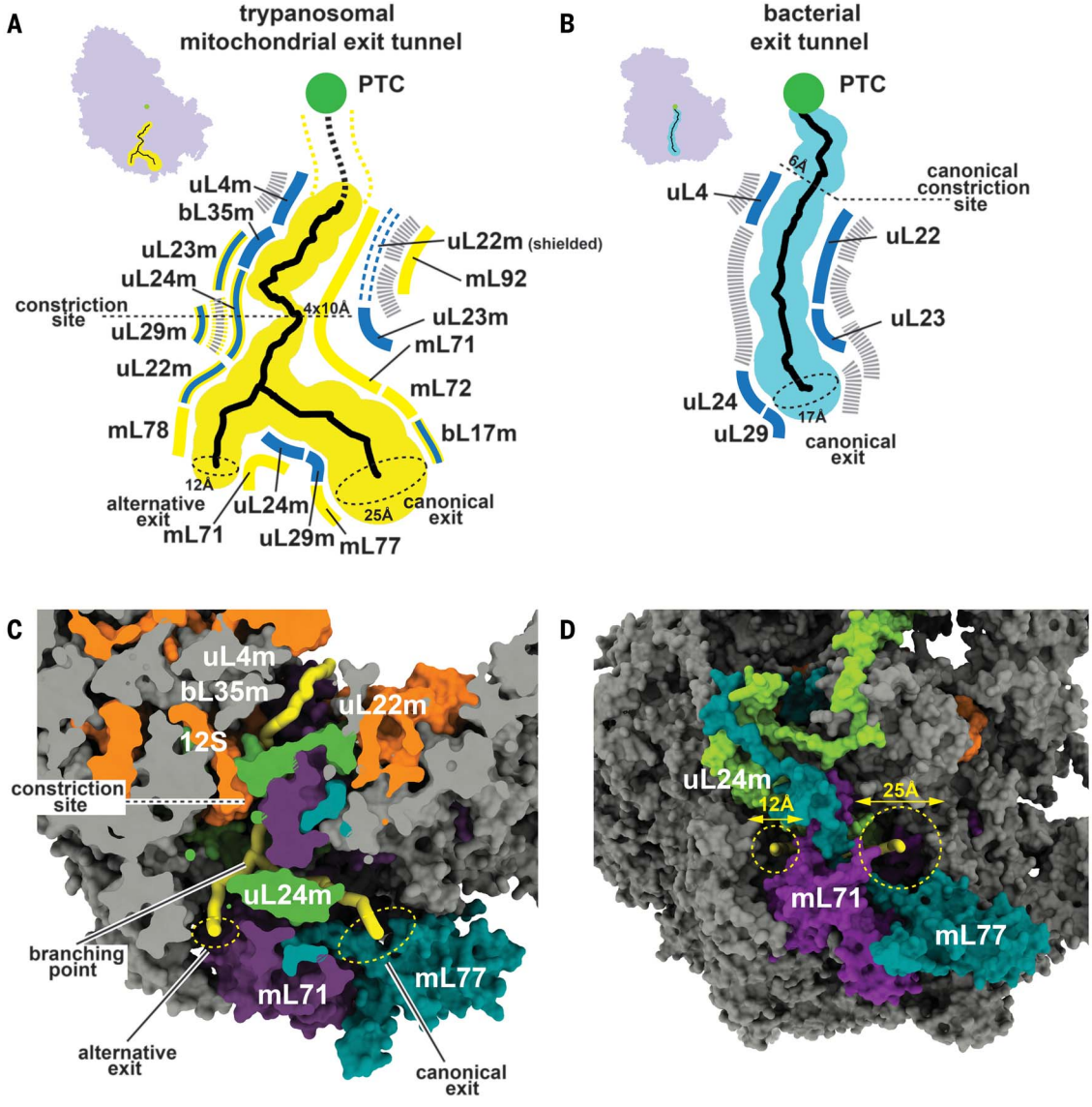


Fig. 5. The trypanosomal mitoribosomal exit tunnel. (**A** and **B**) Schematics of the architecture of the exit tunnel emerging from the PTC (green sphere) found in trypanosomal mitochondria (A) and bacteria (B). The exit tunnels are surrounded by conserved protein (blue lines) and rRNA elements (gray dashed lines). Trypanosomal-specific elements are colored in yellow (proteins), yellow-blue (protein extensions), and gray-yellow (rRNA segment at the tunnel constriction). (**C** and **D**) Surface representation of the tunnel path and the two exit sites (yellow) in a cross section (C) or from the exterior (D), with proteins in gray and the rRNA in orange. The branching of the exit tunnel is mediated by uL24m (green), mL71 (purple), and mL77 (dark cyan).

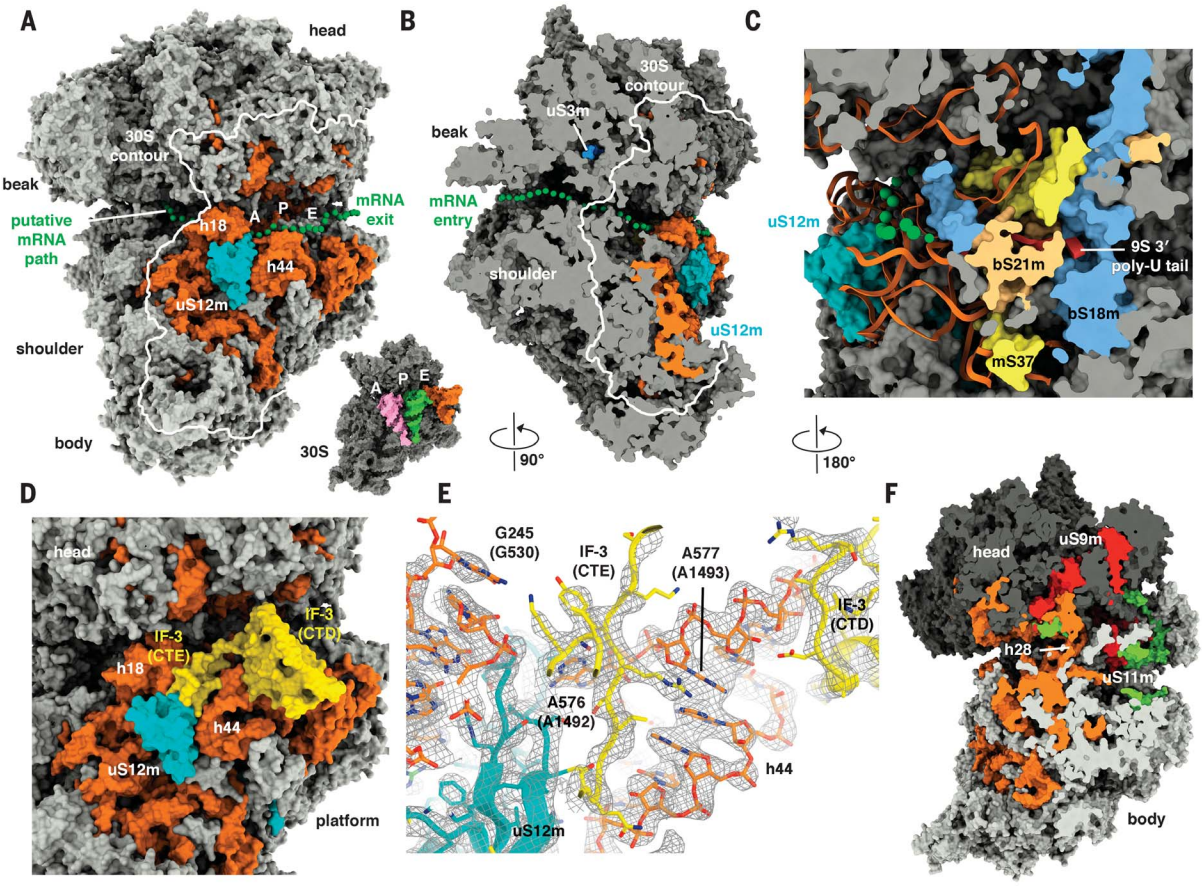


Fig. 6. The small trypanosomal mitoribosomal subunit in complex with IF-3. (**A** to **C**) The putative mRNA path (depicted by green spheres) in the Tb-mt-SSU is shown from the subunit interface and in cross sections through the shoulder (B) or the platform regions (C). Proteins are shown in gray, the 9S rRNA in orange, and uS12m in cyan, together with a contour of the bacterial 30S (white outline). The bacterial 30S subunit in complex with A-, P-, and E-site tRNAs is shown in the inset. Trypanosomal proteins bS18m (blue), bS21m (sand), and mS37 (yellow) partially shield the 3′ poly-U tail (red) from the mRNA path. (**D** and **E**) Interaction of the C-terminal extension (CTE) of the CTD of IF-3 (yellow) with h44 at the decoding center (orange) shown in surface (D) and cartoon (E) representation together with experimental cryo-EM density (gray). (**F**) Cross section of the Tb-mt-SSU showing the head (dark gray) and body (light gray) connection formed by the 9S rRNA (orange) and proteins uS9m (red) and uS11m (green).

mutations of catalytically important amino acids in their active sites, suggesting that they lost their enzymatic activity and rather play an architectural role in the trypanosomal mitoribosome. Furthermore, we identified several bound cofactors that typically participate in enzymatic reactions; whereas in the trypanosomal mitoribosome, they apparently play a purely structural role (fig. S14 and supplementary text). These include nicotinamide adenine dinucleotide that bridges uL4m, uL15m, and mL64, a guanosine triphosphate bound to mS29, and a nucleoside triphosphate (NTP) bound to a positively charged binding pocket created between the helical repeats of mS53, mS57, and the 9*S* rRNA (figs. S12 and S14). Presumably, this NTP represents a relic of the 9*S* rRNA reduction that occurred most prominently in the head region of the Tb-mt-SSU (Fig. 3A).

The remodeled exit tunnel of the large subunit

Because of the recruitment of specific proteins or protein extensions to the trypanosomal mitoribosome, the tunnel and the exit region have evolved unique molecular features. We computationally analyzed the tunnel width to reveal a putative path of the nascent polypeptide (*36*). The trypanosomal tunnel wall is formed mostly by proteins rather than rRNA. In addition to the core folds of conserved proteins uL4m, uL23m, uL24m, and uL29m that still contribute to shaping of the exit tunnel, the tunnel wall is lined by numerous trypanosomal-specific proteins or protein extensions (Fig. 5 and fig. S15). The conserved β sheet of uL22m that usually lines the tunnel is covered by the N-terminal extension of mL71 that changes the path of the tunnel. Further down, mL71 together with trypanosomal-specific elements of uL24m and of the 12*S* rRNA narrows the tunnel elliptically to ~10 × 4 Å^2, which would only allow the passage of an unfolded protein. Notably, the tunnel in yeast mitochondria also contains a constriction in this region, which includes uL22m and uL24m, that has been suggested to have a regulatory role (*28*).

Because of trypanosomal-specific elements in uL24m, mL71, and mL77, the tunnel branches, resulting in a smaller exit with a diameter of ~12 Å and a larger exit with a diameter of ~25 Å (Fig. 5A). The larger tunnel exit points into a similar direction as the canonical bacterial exit, whereas the positioning of the alternative smaller exit is close to the one found in yeast mitoribosomes (*28*). This observation provides for an interesting possibility that the canonical exit is used for the synthesis and cotranslational insertion of membrane proteins, whereas the smaller alternative exit may be used for the synthesis of matrix proteins such as uS3m and uS12m without the need to detach the ribosome from the membrane.

The mRNA path of the small subunit

Modeling of the bacterial mRNA based on superposition of the few conserved protein cores and rRNA elements of the decoding site reveals the location of the mRNA channel, with its entrance and exit regions in the small subunit (Fig. 6A) (*37*). In trypanosomes, uS5m represents the only conserved element facing the mRNA at the entry site, as uS4m does not

exist and uS3m is buried under the trypanosomal-specific protein mS59 and an extension of uS10m. Because of the large beak and shoulder regions, the canonical mRNA entry path below the beak is extended by ~150 Å (Fig. 6B).

The mRNA exit region at the Tb-mt-SSU platform is close to the location of the 3′ end of the 9*S* rRNA, which harbors seven uracil residues that are added posttranscriptionally (*38*). The 3′ end of the 9*S* rRNA occupies a similar position relative to the mRNA channel as the anti–Shine-Dalgarno sequence at the 3′ end of the bacterial 16*S* rRNA (Fig. 6C). Nevertheless, the poly-U tail is sequestered by bS18m, bS21m, and mS37 and is therefore not available for base-pairing interactions with mRNAs.

The small subunit in complex with mitochondrial IF-3

The cryo-EM reconstruction of the isolated small subunit reveals that it is in complex with mitochondrial IF-3, whose CTD can be seen bound to the top of the shortened helix h44 (Fig. 6D). This position is similar to observed for bacterial IF-3 (*39*) and suggests that the Tb-mt-SSU•IF-3 complex represents a post-recycling/preinitiation state that precedes mRNA and initiator tRNA binding during translation initiation.

Currently, the role of mitochondrial IF-3 is not well understood, and our Tb-mt-SSU•IF-3 complex indicates a possible mechanistic role for its C-terminal extension during initiation in trypanosomes. This region of the mitochondrial IF-3 interacts with the decoding center and stabilizes A576 (A1492 in *E. coli*) in a flipped-out position, pointing away from the putative mRNA channel to allow mRNA positioning (Fig. 6E). This function is similar to what was observed for the bacterial IF-1 that interacts with the decoding center and stabilizes bases A1492 and A1493 in a flipped-out conformation (*39*, *40*). This observation explains how the C-terminal extension of trypanosomal mitochondrial IF-3 can compensate for the missing IF-1, which is universally absent in mitochondria (*41*) but is an essential initiation factor in bacteria and yeast (*42*, *43*).

Although the small subunit changed shape due to additions of many new proteins, there is still a well-defined separation between the head and the body to allow for the rotation that is mechanistically important during initiation and for repositioning of tRNAs during translation elongation (Fig. 6F) (*44*). Nevertheless, it is likely that the flexibility between the head and the body is reduced in the trypanosomal small subunit because the two parts are covalently connected not only through rRNA, as is typically the case, but also through several proteins (Fig. 6E). As a consequence, the comparison between the free and complexed small subunits reveals head to body tilting instead of the canonical twisting rotation (fig. S16).

Conclusions

Our results define the composition and the structure of the extremely remodeled mitoribosome of trypanosomes. Its unusual architecture markedly deviates from the structures of all other ribosomes and allows us to better define the universally conserved core of the ribosome that is responsible for the most basic ribosomal functions. We also observe that the architectural role of the rRNA during ribosomal assembly has been taken over by proteins. This implies that ribosomal biogenesis is unlikely to proceed through initial folding of rRNA helices that coalesce into an rRNA dominated tertiary structure with the help of ribosomal proteins and maturation factors. Instead, the trypanosomal ribosome is more likely to assemble through binding of the unfolded rRNA to the preformed elements of the protein shell. The structure also extends our understanding of mitochondrial translation and of the role of mitochondrial IF-3 in particular. Last, because some trypanosomes are human pathogens, this unique structure may be helpful for developing antibiotics to treat sleeping sickness, Chagas disease, or Leishmaniasis.

Materials and methods summary

Purification of the T. brucei *mitoribosome and cryo-EM analysis*

For the affinity purification of a *T. brucei* mitoribosomal sample, we introduced PTP affinity tags at the C-termini of different mitoribosomal proteins (*45*). Such affinity purified complexes were analyzed by electron microscopy and mass spectrometry (*15*).

An aliquot of purified sample was loaded on a holey carbon copper grid that was pre-coated with a thin carbon film. The grid was subsequently plunge frozen using an FEI Vitrobot Mark IV. Individual cryo-EM data sets obtained from either LSU-tagged or SSU-tagged mitoribosomes were collected using a FEI Titan Krios cryo-transmission electron microscope (TEM) equipped with a FEI Falcon III direct electron detector.

Initial references for the Tb-mt-SSU and Tb-mt-LSU were calculated ab initio with EMAN (*46*) using subsets of characteristic 2D class averages from negative stain EM data sets that were either depleted (Tb-mt-SSU) or enriched (Tb-mt-LSU) under dissociating conditions in a mitoribosomal sample carrying a LSU-tag (fig. S2A). These initial models were used for the first steps of particle picking and initial 3D alignments and iteratively improved during the 3D classification process. The structures were refined using masks for the entire Tb-mt-SSU•IF-3, the Tb-mt-SSU head or the Tb-mt-SSU•IF-3 body, the Tb-mt-LSU or without a mask in the case of the complete trypanosomal mitoribosome (fig. S2B). Final data sets included 101,308 particle images for the Tb-mt-SSU head, 31,911 particle images for the Tb-mt-SSU•IF-3 body, 31,619 particle images for the Tb-mt-LSU and 7141 particle images for the complete trypanosomal mitoribosome. Using the FSC = 0.143 criterion, these reconstructions were resolved at 3.1 Å for the Tb-mt-SSU head, 3.3 Å for the Tb-mt-SSU•IF-3 body, 3.4 Å for the Tb-mt-LSU and 7.8 Å for the complete trypanosomal mitoribosome (fig. S2C).

Structure building and model refinement

The atomic model of the trypanosomal ribosome was built into the high-resolution maps of the Tb-mt-SSU head, Tb-mt-SSU•IF-3 body and the Tb-mt-LSU (fig. S3 and table S1) using *Coot* (*47*) and O (*48*, *49*). Only the core areas of the conserved ribosomal homologs were found to share homology with the mammalian mitoribosome or the bacterial ribosome (*5*, *6*, *19*). For most proteins, their fold and identity was deduced from the experimental cryo-EM density by predicting their amino acid side chains according to the density features and their chemical environment followed by sequence homology search.

The 12*S* and 16*S* rRNAs from the mammalian mitoribosome (*5*, *6*) were docked into the structurally conserved core areas of trypanosomal rRNA and served as initial models to establish the correct registry. Additional elements were added to these rRNA models in agreement with the remaining 9*S* and 12*S* rRNA sequences. Whereas for the 9*S* rRNA a complete atomic model was built, we were able to interpret about 50% of the 12*S* rRNA in well-ordered areas. To display the missing parts of the 12*S* rRNA that included regions around the PTC (Figs. 1 to 3), we generated a homology model based on the mammalian mitoribosomal 16*S* rRNA using the predicted secondary structure as a guide (*9*). These parts are not included in the final atomic model.

The atomic models of the Tb-mt-SSU head, Tb-mt-SSU•IF-3 body, the complete Tb-mt-SSU•IF-3 and the Tb-mt-LSU were refined with PHENIX (*50*, *51*) using a combination of secondary structure restrained real and reciprocal space refinement. F_{obs}–F_{calc} difference Fourier maps were used to detect and eliminate errors and to identify missing co-factors and additional ions. Table S1 summarizes the refinement and provides model statistics. For the model of the complete mitoribosome, the high-resolution coordinates were docked into the 7.8 Å cryo-EM map of the complete trypanosomal mitoribosome.

REFERENCES AND NOTES

1. T. M. Schmeing, V. Ramakrishnan, What recent ribosome structures have revealed about the mechanism of translation. *Nature* **461**, 1234–1242 (2009). doi: 10.1038/nature08403 pmid: 19838167
2. R. R. Gutell, B. Weiser, C. R. Woese, H. F. Noller, Comparative anatomy of 16-S-like ribosomal RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **32**, 155–216 (1985). doi: 10.1016/S0079-6603(08)60348-7 pmid: 3911275
3. C. R. Woese, G. E. Fox, Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5088–5090 (1977). doi: 10.1073/pnas.74.11.5088 pmid: 270744
4. S. Melnikov *et al.*, One core, two shells: Bacterial and eukaryotic ribosomes. *Nat. Struct. Mol. Biol.* **19**, 560–567 (2012). doi: 10.1038/nsmb.2313; pmid: 22664983
5. A. Amunts, A. Brown, J. Toots, S. H. W. Scheres, V. Ramakrishnan, The structure of the human mitochondrial ribosome. *Science* **348**, 95–98 (2015). doi: 10.1126/science.aaa1193 pmid: 25838379

6. B. J. Greber *et al*., The complete structure of the 55S mammalian mitochondrial ribosome. *Science* **348**, 303–308 (2015). doi: 10.1126/science.aaa3872; pmid: 25837512
7. V. F. de la Cruz, J. A. Lake, A. M. Simpson, L. Simpson, A minimal ribosomal RNA: Sequence and secondary structure of the 9S kinetoplast ribosomal RNA from Leishmania tarentolae. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1401–1405 (1985). doi: 10.1073/pnas.82.5.1401; pmid: 3856267
8. I. C. Eperon, J. W. Janssen, J. H. Hoeijmakers, P. Borst, The major transcripts of the kinetoplast DNA of Trypanosoma brucei are very small ribosomal RNAs. *Nucleic Acids Res.* **11**, 105–125 (1983). doi: 10.1093/nar/11.1.105; pmid: 6306559
9. P. Sloof *et al*., Further characterization of the extremely small mitochondrial ribosomal RNAs from trypanosomes: A detailed comparison of the 9S and 12S RNAs from Crithidia fasciculata and Trypanosoma brucei with rRNAs from other organisms. *Nucleic Acids Res.* **13**, 4171–4190 (1985). doi: 10.1093/nar/13.11.4171; pmid: 2409531
10. D. A. Maslov *et al*., Isolation and characterization of mitochondrial ribosomes and ribosomal subunits from *Leishmania tarentolae*. *Mol. Biochem. Parasitol.* **148**, 69–78 (2006). doi: 10.1016/j.molbiopara.2006.02.021; pmid: 16600399
11. H. H. Shu, H. U. Göringer, Trypanosoma brucei mitochondrial ribonucleoprotein complexes which contain 12S and 9S ribosomal RNAs. *Parasitology* **116**, 157–164 (1998). doi: 10.1017/S0031182097002023; pmid: 9509025
12. A. Zíková *et al*., *Trypanosoma brucei* mitochondrial ribosomes: Affinity purification and component identification by mass spectrometry. *Mol. Cell. Proteomics* **7**, 1286–1296 (2008). doi: 10.1074/mcp.M700490-MCP200; pmid: 18364347
13. N. Desai, A. Brown, A. Amunts, V. Ramakrishnan, The structure of the yeast mitochondrial ribosome. *Science* **355**, 528–531 (2017). doi: 10.1126/science.aal2415; pmid: 28154081
14. S. Klinge, F. Voigts-Hoffmann, M. Leibundgut, N. Ban, Atomic structures of the eukaryotic ribosome. *Trends Biochem. Sci.* **37**, 189–198 (2012). doi: 10.1016/j.tibs.2012.02.007; pmid: 22436288
15. Materials and methods are available as supplementary materials.
16. P. Bieri, B. J. Greber, N. Ban, High-resolution structures of mitochondrial ribosomes and their functional implications. *Curr. Opin. Struct. Biol.* **49**, 44–53 (2018). doi: 10.1016/j.sbi.2017.12.009; pmid: 29348055
17. M. R. Sharma, T. M. Booth, L. Simpson, D. A. Maslov, R. K. Agrawal, Structure of a mitochondrial ribosome with minimal RNA. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 9637–9642 (2009). doi: 10.1073/pnas.0901631106; pmid: 19497863
18. D.-E. Mbang-Benet *et al*., RNA interference screen reveals a high proportion of mitochondrial proteins essential for correct cell cycle progress in *Trypanosoma brucei*. *BMC Genomics* **16**, 297 (2015). doi: 10.1186/s12864-015-1505-5 pmid: 25888089
19. M. Selmer *et al*., Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935–1942 (2006). doi: 10.1126/science.1131127; pmid: 16959973
20. C. Marchal, N. Ismaïli, E. Pays, A ribosomal S12-like gene of *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **57**, 331–334 (1993). doi: 10.1016/0166-6851(93)90208-F; pmid: 8433721
21. I. Aphasizheva, D. A. Maslov, R. Aphasizhev, Kinetoplast DNA-encoded ribosomal protein S12: A possible functional link between mitochondrial RNA editing and translation in Trypanosoma brucei. *RNA Biol.* **10**, 1679–1688 (2013). doi: 10.4161/rna.26733; pmid: 24270388
22. A. B. Matassova, M. V. Rodnina, W. Wintermeyer, Elongation factor G-induced structural change in helix 34 of 16S rRNA related to translocation on the ribosome. *RNA* **7**, 1879–1885 (2001). pmid: 11780642
23. A. H. Ratje *et al*., Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. *Nature* **468**, 713–716 (2010). doi: 10.1038/nature09547; pmid: 21124459
24. Ü. Allas, A. Liiv, J. Remme, Functional interaction between RNase III and the Escherichia coli ribosome. *BMC Mol. Biol.* **4**, 8 (2003). doi: 10.1186/1471-2199-4-8; pmid: 12814522
25. S. Mohan, H. F. Noller, Recurring RNA structural motifs underlie the mechanics of L1 stalk movement. *Nat. Commun.* **8**, 14285 (2017). doi: 10.1038/ncomms14285; pmid: 28176782
26. L. G. Trabuco *et al*., The role of L1 stalk–tRNA interaction in the ribosome elongation cycle. *J. Mol. Biol.* **402**, 741–760 (2010). doi: 10.1016/j.jmb.2010.07.056; pmid: 20691699
27. A. Brown *et al*., Structure of the large ribosomal subunit from human mitochondria. *Science* **346**, 718–722 (2014). doi: 10.1126/science.1258026; pmid: 25278503
28. B. J. Greber *et al*., The complete structure of the large subunit of the mammalian mitochondrial ribosome. *Nature* **515**, 283–286 (2014). doi: 10.1038/nature13895; pmid: 25271403
29. A. Amunts *et al*., Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485–1489 (2014). doi: 10.1126/science.1249410; pmid: 24675956
30. G. S. Groot, T. L. Mason, N. Van Harten-Loosbroek, *Var*$_1$ is associated with the small ribosomal subunit of mitochondrial ribosomes in yeast. *Mol. Gen. Genet.* **174**, 339–342 (1979). doi: 10.1007/BF00267808; pmid: 384172
31. J. Kyte, R. F. Doolittle, A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982). doi: 10.1016/0022-2836(82)90515-0; pmid: 7108955
32. M. A. Andrade, C. Petosa, S. I. O'Donoghue, C. W. Müller, P. Bork, Comparison of ARM and HEAT protein repeats. *J. Mol. Biol.* **309**, 1–18 (2001). doi: 10.1006/jmbi.2001.4624; pmid: 11491282
33. S. Manna, An overview of pentatricopeptide repeat proteins and their applications. *Biochimie* **113**, 93–99 (2015). doi: 10.1016/j.biochi.2015.04.004; pmid: 25882680
34. I. Aphasizheva *et al*., Ribosome-associated pentatricopeptide repeat proteins function as translational activators in mitochondria of trypanosomes. *Mol. Microbiol.* **99**, 1043–1058 (2016). doi: 10.1111/mmi.13287; pmid: 26713541
35. M. Pusnik, I. Small, L. K. Read, T. Fabbro, A. Schneider, Pentatricopeptide repeat proteins in *Trypanosoma brucei* function in mitochondrial ribosomes. *Mol. Cell. Biol.* **27**, 6876–6888 (2007). doi: 10.1128/MCB.00708-07; pmid: 17646387
36. K. Berka *et al*., MOLE*online* 2.0: Interactive web-based analysis of biomacromolecular channels. *Nucleic Acids Res.* **40**, W222–W227 (2012). doi: 10.1093/nar/gks363; pmid: 22553366
37. L. B. Jenner, N. Demeshkina, G. Yusupova, M. Yusupov, Structural aspects of messenger RNA reading frame maintenance by the ribosome. *Nat. Struct. Mol. Biol.* **17**, 555–560 (2010). doi: 10.1038/nsmb.1790; pmid: 20400952
38. B. K. Adler, M. E. Harris, K. I. Bertrand, S. L. Hajduk, Modification of Trypanosoma brucei mitochondrial rRNA by posttranscriptional 3′ polyuridine tail formation. *Mol. Cell. Biol.* **11**, 5878–5884 (1991). doi: 10.1128/MCB.11.12.5878; pmid: 1719373
39. T. Hussain, J. L. Llácer, B. T. Wimberly, J. S. Kieft, V. Ramakrishnan, Large-scale movements of IF3 and tRNA during bacterial translation initiation. *Cell* **167**, 133–144.e13 (2016). doi: 10.1016/j.cell.2016.08.074; pmid: 27662086
40. A. P. Carter *et al*., Crystal structure of an initiation factor bound to the 30S ribosomal subunit. *Science* **291**, 498–501 (2001). doi: 10.1126/science.1057766; pmid: 11228145
41. G. C. Atkinson *et al*., Evolutionary and genetic analyses of mitochondrial translation initiation factors identify the missing mitochondrial IF3 in *S. cerevisiae*. *Nucleic Acids Res.* **40**, 6122–6134 (2012). doi: 10.1093/nar/gks272; pmid: 22457064
42. H. S. Cummings, J. W. Hershey, Translation initiation factor IF1 is essential for cell viability in Escherichia coli. *J. Bacteriol.* **176**, 198–205 (1994). doi: 10.1128/jb.176.1.198-205.1994; pmid: 8282696
43. C.-L. Wei, M. Kainuma, J. W. B. Hershey, Characterization of yeast translation initiation factor 1A and cloning of its essential gene. *J. Biol. Chem.* **270**, 22788–22794 (1995). doi: 10.1074/jbc.270.39.22788; pmid: 7559407
44. S. Mohan, J. P. Donohue, H. F. Noller, Molecular mechanics of 30S subunit head rotation. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 13325–13330 (2014). doi: 10.1073/pnas.1413731111; pmid: 25187561
45. B. Schimanski, T. N. Nguyen, A. Günzl, Highly efficient tandem affinity purification of trypanosome protein complexes based on a novel epitope combination. *Eukaryot. Cell* **4**, 1942–1950 (2005). doi: 10.1128/EC.4.11.1942-1950.2005; pmid: 16278461
46. S. J. Ludtke, P. R. Baldwin, W. Chiu, EMAN: Semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999). doi: 10.1006/jsbi.1999.4174; pmid: 10600563
47. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of *Coot*. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010). doi: 10.1107/S0907444910007493; pmid: 20383002
48. T. A. Jones, Interactive electron-density map interpretation: From *INTER* to *O*. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2115–2125 (2004). doi: 10.1107/S0907444904023509; pmid: 15572764
49. T. A. Jones, J. Y. Zou, S. W. Cowan, M. Kjeldgaard, Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991). doi: 10.1107/S0108767390010224; pmid: 2025413
50. P. D. Adams *et al*., *PHENIX*: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010). doi: 10.1107/S0907444909052925; pmid: 20124702
51. P. V. Afonine *et al*., Towards automated crystallographic structure refinement with *phenix.refine*. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012). doi: 10.1107/S0907444912001308; pmid: 22505256

ACKNOWLEDGMENTS

We thank E. Kummer, A. Scaiola, A. Jomaa, C. H. S. Aylett, B. Schimanski, and C. M. T. Spahn for support and discussions. Cryo-EM data were collected at the Scientific Center for Optical and Electron Microscopy (ScopeM) of ETH Zurich, and we thank P. Tittmann (ScopeM) for the support. A.L. would like to thank R. Aebersold for access to instrumentation and infrastructure. The Orbitrap Elite mass spectrometer was funded by an ETH Scientific Equipment grant. The cryo-EM data were processed on the Euler computing cluster of the Swiss National Supercomputing Centre (CSCS). **Funding:** This work was supported by the Swiss National Science Foundation (SNSF) and the National Center of Excellence in Research (NCCR) RNA and disease program of the SNSF. D.J.F.R. was supported by a FEBS long-term fellowship. **Author contributions:** D.J.F.R., M.N., A.S., and N.B. designed the experiment. M.N. and E.K.H. cloned the *T. brucei* strains. M.N., E.K.H., D.J.F.R., and C.P. purified the mitochondrial ribosomes. A.L. designed and performed the mass spectrometry analysis. D.J.F.R. and C.P. prepared cryo-EM grids. D.J.F.R. and D.B. acquired the cryo-EM data. D.J.F.R. and D.B. calculated the cryo-EM reconstructions. D.J.F.R., M.L., P.B., D.B., and N.B. interpreted the structures. D.J.F.R., M.L., P.B., and N.B. wrote the manuscript. All authors contributed to the final version of the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The cryo-EM maps and PDB coordinates have been deposited in the Electron Microscopy Databank and PDB database, respectively. These entries include the complex of the complete *T. brucei* mitoribosome (EMD-0229 and PDB-6HIV), the complete Tb-mt-SSU•IF-3 (EMD-0230 and PDB-6HIW), the Tb-mt-LSU (EMD-0231 and PDB-6HIX), the Tb-mt-SSU•IF-3 body (EMD-0232 and PDB-6HIY), and the Tb-mt-SSU head (EMD-0233 and PDB-6HIZ). A PyMOL script for display of the complete trypanosomal mitoribosome is available from the Ban Lab website (www.bangroup.ethz.ch/).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6413/eaau7735/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S16
Tables S1 and S2
References (*52*–*93*)
Data S1

13 July 2018; accepted 3 September 2018
Published online 13 September 2018
10.1126/science.aau7735

RESEARCH ARTICLE

NEUROSCIENCE

Dynamic salience processing in paraventricular thalamus gates associative learning

Yingjie Zhu[1,2], Gregory Nachtrab[1], Piper C. Keyes[1,3], William E. Allen[1,3], Liqun Luo[1,4], Xiaoke Chen[1]*

The salience of behaviorally relevant stimuli is dynamic and influenced by internal state and external environment. Monitoring such changes is critical for effective learning and flexible behavior, but the neuronal substrate for tracking the dynamics of stimulus salience is obscure. We found that neurons in the paraventricular thalamus (PVT) are robustly activated by a variety of behaviorally relevant events, including novel ("unfamiliar") stimuli, reinforcing stimuli and their predicting cues, as well as omission of the expected reward. PVT responses are scaled with stimulus intensity and modulated by changes in homeostatic state or behavioral context. Inhibition of the PVT responses suppresses appetitive or aversive associative learning and reward extinction. Our findings demonstrate that the PVT gates associative learning by providing a dynamic representation of stimulus salience.

The brain constantly receives streams of complex sensory inputs and must direct attention to the most important or salient stimulus. The salience of a stimulus is determined by both physical properties and behavioral relevance, such as the reward value or novelty (*1*–*5*). Although physical properties, such as brightness or color, are fixed attributes of the stimulus, the behavioral relevance is a relative property that depends on past experience, current homeostatic state, and behavioral context (*1*, *2*, *5*). Therefore, identifying the essential anatomical substrates for tracking the dynamics of stimulus' behavioral relevance is necessary to understand the neural mechanisms underlying proper allocation of attentional resources and to directly examine the contribution of stimulus salience to learning (*6*–*8*).

Early studies largely focused on cortical circuitry and identified the frontoparietal attention network for attentional selection of behaviorally relevant stimuli (*2*, *9*–*11*). Recent work has begun to reveal thalamic contributions to the persistence of frontal cortical activity during motor preparation, working memory, and rule representation (*12*–*15*). However, the coding of various forms of behavioral relevance in the thalamus has not been systematically studied. It remains unclear whether the thalamus can represent context-dependent dynamics of behavioral relevance. If so, it will be important to determine how salience responses in the thalamus contribute to associative learning.

The thalamus is composed of several anatomically and functionally distinct subnuclei. Among them, the paraventricular thalamus (PVT) is particularly situated for integrating information applicable to behavioral relevance (*16*–*24*). The PVT is not directly connected with sensory cortices but is reciprocally connected with regions involved in top-down control, such as the prefrontal and insular cortices. The PVT also receives extensive inputs from the hypothalamus and brainstem, which convey signals about motivational arousal and homeostatic states. In turn, the PVT is the only thalamic nucleus that innervates all structures in the extended amygdala system (*16*, *22*, *24*–*26*). Previous lesion and pharmacological silencing studies suggested potential roles of the PVT in both appetitive and defensive behaviors. However, little is known about how PVT neurons engage in behaviors with opposite valence.

PVT encodes multiple forms of salience

We infected PVT neurons (between Bregma −1.06 to −1.58 mm) with adeno-associated virus (AAV) expressing a genetically encoded Ca^{2+} indicator (AAV-GCaMP6m) (*27*) then used fiber photometry to record population Ca^{2+} signals in the PVT of the head-fixed mice across days of associative learning (fig. S1) (*28*–*30*). We first randomly presented water-restricted mice with odors or water (5 μl) without pairing them (materials and methods) (*30*). Initially, both stimuli robustly activated the PVT, but whereas PVT responses to free water remained consistent, odor-evoked responses were rapidly diminished (Fig. 1A). Habituation of PVT responses to odor was stimulus-specific and long-lasting because subsequent exposure to a different odor still elicited robust PVT responses (fig. S2A), and PVT response to the same odor was still strongly suppressed 2 days later (fig. S2B). Similar novelty responses were also observed across multiple modalities, including visual and auditory stimuli, in the PVT (fig. S2C).

After the odors were no longer novel, we then trained the mice to associate the same set of odors with either appetitive, neutral, or aversive outcomes (*17*, *31*). Each training trial began with a conditioned stimulus (CS) (1 s odor), followed by a 2-s delay and an unconditioned stimulus (US) (the outcome) (Fig. 1B). As training progressed, mice began to display anticipatory licks only during the delay of appetitive (water reward) trials, indicating the establishment of CS-US association (Fig. 1C and fig. S3A) (*17*, *31*). The familiar odors gradually gained behavioral relevance as they were associated with reinforcing outcomes. After the mice had fully learnt the task, we performed fiber photometry recording and observed robust task-evoked responses in the PVT of GCaMP6- but not enhanced green fluorescent protein (eGFP)–expressing mice (Fig. 1D and fig. S3B). The PVT responded to both CS and US irrespective of appetitive or aversive outcomes (Fig. 1D), and their averaged response magnitudes were graded, reflecting the intensity of reward (5 versus 15 μl water) and punishment (air puff versus tail shock) (Fig. 1, F and G).

Because fiber photometry records population activities, it is still possible that within the PVT, subpopulations might encode a specific valence. We therefore performed in vivo single-unit recording and found that a majority of recorded PVT neurons (85 of 115) were responsive to the learned task. Among them, 68% of task-related neurons were excited by CS or US of both appetitive and aversive outcomes, a hallmark of salience coding (Fig. 1E) (*32*, *33*). The other 32% of neurons can encode valence because they responded heterogeneously to the appetitive versus aversive task (Fig. 1E). Although cross-correlation analysis and lick-triggered spike analysis revealed little correlation between licking with action potential firing in the PVT (fig. S4), the timing of CS responses in appetitive trials was more distributed and tiled the entire delay period (Fig. 1E). This suggests that the PVT activity encodes salience of both CS and US and can reflect the level of behavioral engagement.

Prediction error (PE) signals the discrepancy between the expected and actual received outcomes (*6*). It is encoded by a widespread neuronal network and provides teaching signals for associative learning (*31*, *34*). To test whether the PVT can encode PE, we chronically recorded the PVT across days of associative learning, pairing cues with a water reward or an air puff. Following behavioral training, CS responses in the PVT gradually increased, whereas US responses remained constant in both appetitive and aversive

[1]Department of Biology, Stanford University, Stanford, CA 94305, USA. [2]Guangdong Provincial Key Laboratory of Brain Connectome and Behavior, The Brain Cognition and Brain Disease Institute, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China. [3]Neurosciences Program, Stanford University, Stanford, CA 94305, USA. [4]Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA.
*Corresponding author. Email: xkchen@stanford.edu

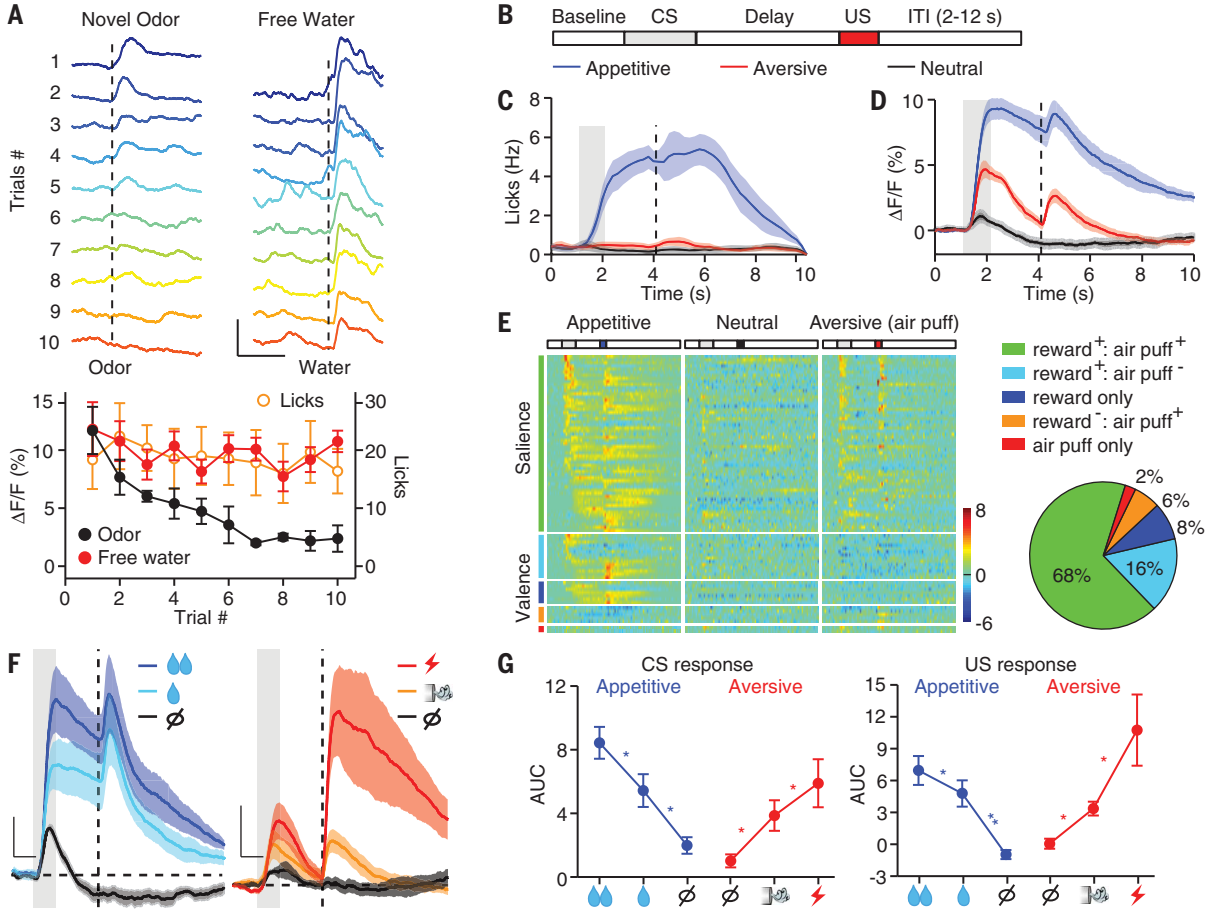


Fig. 1. PVT neurons encode salience irrespective of valence. (**A**) (Top) Photometric traces of calcium responses in the PVT to 10 repetitions of randomized (left) odor and (right) water (5 μl) stimuli. Dashed line indicates the time of stimulus delivery. Scale bar, 10% change in fluorescence intensity (Δ*F*/*F*), 3 s. (Bottom) Left *y* axis, quantification of odor (black dot) and free water (red dot) evoked Δ*F*/*F* over 10 repetitions. Right *y* axis, quantification of free water (orange circle) evoked licks over 10 repetitions. Novel odor, *n* = 6 mice; free water, *n* = 12 mice; licks, *n* = 12 mice. (**B**) Trial structure of the Pavlovian conditioning paradigm. ITI, intertrial interval; CS, conditioned stimulus; US, unconditioned stimulus. (**C**) Mean lick rate of well-trained animals (*n* = 7 mice) shows anticipatory licking in appetitive (blue) but not neutral (black) and aversive (red) trials. Gray bar indicates 1 s of CS delivery; vertical dashed line indicates US delivery. (**D**) Mean photometric responses of the PVT to CS and US in both appetitive (blue) and aversive (red) but not neutral (black) trials. Shade, SEM across mice; *n* = 7 mice. (**E**) *Z* score heat maps (left) and pie chart (right) for all task-responding neurons identified by means of in vivo single-unit recording during Pavlovian tasks of well-trained animals. Neurons are separated into five subgroups on the basis of their tuning properties and are rank-ordered by their response onset times during reward cue stimulation. Each row in the heat maps represents responses from the same neuron to different stimuli. *n* = 85 neurons from 12 mice. (**F** and **G**) Mean photometric responses (F) (*n* = 7 mice) and quantification (G) showing that CS and US response in the PVT are graded to different intensity of reward (left, 5 versus 15 μl water) and punishment (right, air puff versus tail shock). AUC, area under curve (materials and methods). Scale bars, 2% [(F), left], 4% [(F), right] Δ*F*/*F*, 1 s. Big reward, small reward, and nothing: *n* = 7, 7, and 7 mice, respectively; Big punishment, small punishment, and nothing: *n* = 6, 6, and 5 mice, respectively. **P* < 0.05, ***P* < 0.01 [One-way analysis of variance (ANOVA), post-hoc Tukey's test]. Shade, SEM across mice in (C), (D), and (F). Data are means ± SEM.

learning (Fig. 2, A and B). Moreover, in well-trained animals, the magnitude of PVT responses to a well-predicted US was similar to that of unexpected delivery of a US (Fig. 2, C to F, and fig. S5). Therefore, PVT does not encode PE because PE-encoding neurons gradually decrease their US responses during associative learning, and unexpected events should have bigger US responses.

PVT activity controls associative learning

Salient stimuli attract attention, which in turn facilitates associative learning. If CS- or US-evoked responses in the PVT represent stimulus salience, then suppression of PVT responses should decrease the efficiency of CS-US association. To test this, we infected PVT neurons with AAV expressing archaerhodopsin-3 (AAV-ArchT) or eGFP (AAV-eGFP) as a control (fig. S1) (*35*). We optogenetically inhibited the PVT (in PVT::ArchT mice) during the cue + delay period or after the US delivery and examined its effect on appetitive or aversive learning during both conditioning sessions (from D1 to D5) and the no-laser (NL) test (Fig. 3, A to D) (*17*). We found that PVT inhibition during either CS (Fig. 3, A and B) or US (Fig. 3, C and D) periods during training reduced anticipatory licking in both conditioning sessions and in the NL test. Optogenetic inhibition during the intertrial interval had no effect (fig. S6A), and PVT inhibition during CS + delay period had no effect on anticipatory licking in well-trained mice (Fig. 3E). Moreover, in a go/no-go task (Fig. 3F), PVT inhibition during cue + delay period reduced licking in go trials but increased licking in no-go trials (Fig. 3G and fig. S7), thus decreasing discriminability (Fig. 3H). Together, CS and US responses in the PVT are required for the formation but not the expression of conditioned reward-seeking.

To examine the impact of PVT inhibition on associative aversive learning, we first trained mice with two different odors both associated

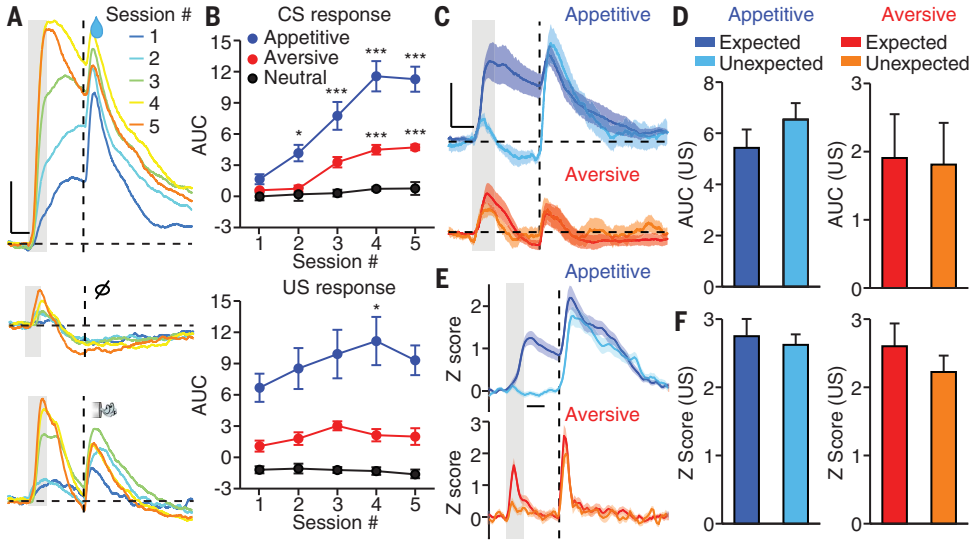


Fig. 2. Dynamics of salience response in PVT neurons during associative learning. (**A**) Representative photometric responses in the PVT across multiple sessions of Pavlovian conditioning (sessions 1 to 5). Gray bar indicates 1 s of CS delivery; vertical dashed line indicates US delivery; and horizontal dashed line indicates baseline Δ*F*/*F*. (Top) Appetitive. (Middle) Nothing. (Bottom) Aversive. Scale bar, 5% Δ*F*/*F*. Each photometric trace is averaged from all 50 trials within a single conditioning session. (**B**) Quantification of CS response (top) and US response (bottom) across five training sessions (*n* = 6 mice). Shown is a significant increase of CS responses after training, whereas the US responses remain consistent; **P* < 0.05, ****P* < 0.001 (two-way ANOVA, post-hoc Bonferroni test). (**C**) Mean photometric responses of the PVT to expected and unexpected delivery of reward (top, *n* = 10 mice; dark blue, expected; light blue, unexpected) and punishment (bottom, *n* = 10 mice; red, expected; orange, unexpected). Scale bar, 4% Δ*F*/*F*. (**D**) Quantification of (C). Wilcoxon signed-rank test; *P* = 0.32 (reward); *P* = 0.49 (punishment). (**E**) Mean *Z* score of single-unit responses of the PVT neurons to expected and unexpected delivery of reward (top, *n* = 31 neurons) and punishment (bottom, *n* = 22 neurons). (**F**) Quantification of (E). Wilcoxon signed-rank test; *P* = 0.28 (reward); *P* = 0.11 (punishment). Gray bar: 1 s of CS delivery; vertical dashed line: US delivery; scale bar: 1 s in (A), (C), and (E). Shade, SEM across mice in (C) and (E). Data are means ± SEM.

with water reward in phase 1, then switched the outcome of odor B to water + tail shock in phase 2 (fig. S8A). The switch caused gradual suppression of odor B–elicited anticipatory licking over the next 5 days of training (fig. S8, B and C). PVT silencing had no effect on nociceptive responses to thermal stimuli on the tail (fig. S8C, inset), but optogenetic inhibition of the PVT (in PVT::ArchT mice) during the cue + delay (fig. S8, B and C) reduced the suppression effect in both conditioning sessions and in the NL test. PVT inhibition during US delivery period had no effect on aversive learning (fig. S8D). Together, these results indicate that PVT activity during the cue period is required for both associative reward and aversive learning and substantiate the critical role of cue salience in driving associative learning.

PVT tracks context-dependent salience

Besides learning, changes of homeostatic state or external environment also influence the perceived salience of sensory stimuli (*2*, *7*, *10*, *36*). If PVT activity represents the salience of CS and US, then their evoked activity in the PVT should also be modulated by internal and external factors. Because we used water as the reward during Pavlovian conditioning, we examined the impact of thirsty versus sated state on CS- and US-evoked PVT activity. The water-predicting cue elicited robust anticipatory licking in well-trained thirsty mice. We then gave these mice free access to 0.6 ml of water to drink until sated. As expected, the same odor cue no longer elicited anticipatory licking in sated mice (Fig. 4A). Using fiber photometry, we recorded PVT activity in both thirsty and sated states and found that both CS- and US-evoked PVT activity were strongly suppressed in sated mice, which is consistent with a decrease of salience of both the water-predicting cue and of water consumption in sated mice (Fig. 4, B and C). A stronger PVT response to air puff was observed in sated than thirsty mice, indicating that an air puff became more salient when homeostatic needs were met and supporting the hypothesis that PVT activity represents context-dependent evaluation of salience between different sensory stimuli.

To further test this hypothesis, we manipulated the behavioral context by changing the intensity of the aversive stimuli and examined the impact of this change on reward responses in the PVT. We first conditioned the mice for 5 days in a mild aversion context, in which an air puff was used as punishment. On day 6, we switched the punishment from air puff to tail shock (strong aversion context) (Fig. 5A). Switching from a mild to a strong aversion context rapidly suppressed reward responses in the PVT, as revealed with both in vivo calcium imaging and single-cell electrophysiological recording (Fig. 5, B to G). Among task-related responders in the PVT, only 76% responded to the aversive condition in the mild aversion context compared with 97% in the strong aversion context, whereas only 80% responded to the reward condition in the latter context compared with 92% in the former ($P < 0.001$, χ^2 test) (Figs. 1E and 5, E and F). This observation revealed the reallocation of salience from appetitive to aversive stimuli after switching from a mild to a strong aversion context, which is consistent with the notion that individual PVT neurons are tuned to the salience of the sensory stimuli, irrespective of its valence. Moreover, the learning rate for the cue-reward association was slower in the strong aversive context (Fig. 5H), which is consistent with the finding that smaller PVT responses were allocated to reward in the strong aversive context than that in the mild aversive context, further supporting that sensory stimuli–evoked PVT response controls the efficiency of associative learning. Together, PVT activity represents the dynamics of stimulus salience after a change of the behavioral context or homeostatic state.

Reward omission responses in the PVT

The above studied salience responses were all evoked by sensory stimulus. Could PVT activity also represent an emotionally salient state without a sensory stimulus? We thus examined reward omission response in the PVT (*18*, *37*). Although no sensory stimulus is delivered, omission of an expected reward is behaviorally relevant. In well-trained mice, we omitted the predicted water reward in a random 10% of trials and recorded PVT activity. Because the CS was delivered before omission, the CS evoked similar PVT responses in both reward trials and reward-omission trials (Fig. 6, A and B). The PVT responded very differently to reward delivery and reward omission. Reward omission lacked the immediate response that was previously observed with water consumption and instead elicited a delayed long-lasting response in the PVT (Fig. 6, A and B). Two possible scenarios might underlie omission responses in the PVT: One might reflect a cognitive state of expectant waiting because many PVT neurons show an anticipatory response to reward delivery during the delay period between CS and US; the other explanation is that the lack of expected water is a salient stimulus, thus activating the PVT (*18*, *37*). We noticed that omission trial responses generally occurred after licking stopped, and when we aligned the calcium response to the

last lick, we observed a rapid increase in PVT activity after the cessation of licking (Fig. 6C). Together, these results suggested that PVT activity could also represent a behaviorally relevant state without sensory stimulus.

Animals use the outcome information of their previous choice to adjust subsequent expectations. How might the reward-omission response in the PVT contribute to this behavioral adjustment? Continuous reward omission will extinguish the learned association. Thus, PVT responses to reward omission might serve as a teaching signal for extinction. To test this directly, we first conditioned PVT::ArchT and PVT::GFP mice with odor cue and water reward for 5 days and examined the cued reward-seeking behavior for the first 10 trials on day 6. We then optogenetically silenced the PVT during the reward omission window in the following extinction trials. Stopping reward delivery caused rapid extinction of cue-evoked anticipatory licking in PVT::GFP mice, whereas the rate of extinction was significantly slower in PVT::ArchT mice (Fig. 6D). Because extinction is also a form of learning, the CS response in the PVT should also be important for extinction. Inhibiting the PVT during the cue + delay period did significantly slow the rate of extinction (Fig. 6E). This suggests that the function of the PVT CS response is to maintain the salience of the CS, which allows for effective learning if the US is changed. These data, together with results in Fig. 3 and fig. S8, substantiate that salience activity in the PVT controls the rate of multiple forms of associative learning.

Discussion

Here, we show that PVT neurons encode multiple salient features of sensory stimuli, including reward, aversion, novelty, and surprise (reward omission). We further demonstrate that PVT provides dynamic representation of salience by manipulating behavioral relevance of stimuli through associative learning, modulation of homeostatic states, and alterations of the behavioral context. When animals are in the behavior context with a mild aversive stimulus, the majority of responders in the PVT respond to appetitive stimuli (Fig. 1E); but when the aversive stimulus is strong, almost all PVT responders respond to aversive stimuli (Fig. 5F). This finding demonstrates that individual PVT neurons track the context-dependent dynamics of salience information. These results also suggest that the PVT has a more specific role than promoting general arousal because PVT reward responses are decreased in the strong aversive context when the animal should be more aroused. How do PVT neurons acquire such response flexibility? Because most thalamic neurons do not have local excitatory connections, we anticipate that PVT inputs play important roles. Further work is required to silence each individual input while examining salience responses in the PVT.

US responses in the PVT are not suppressed by expectation, and reward omission activates

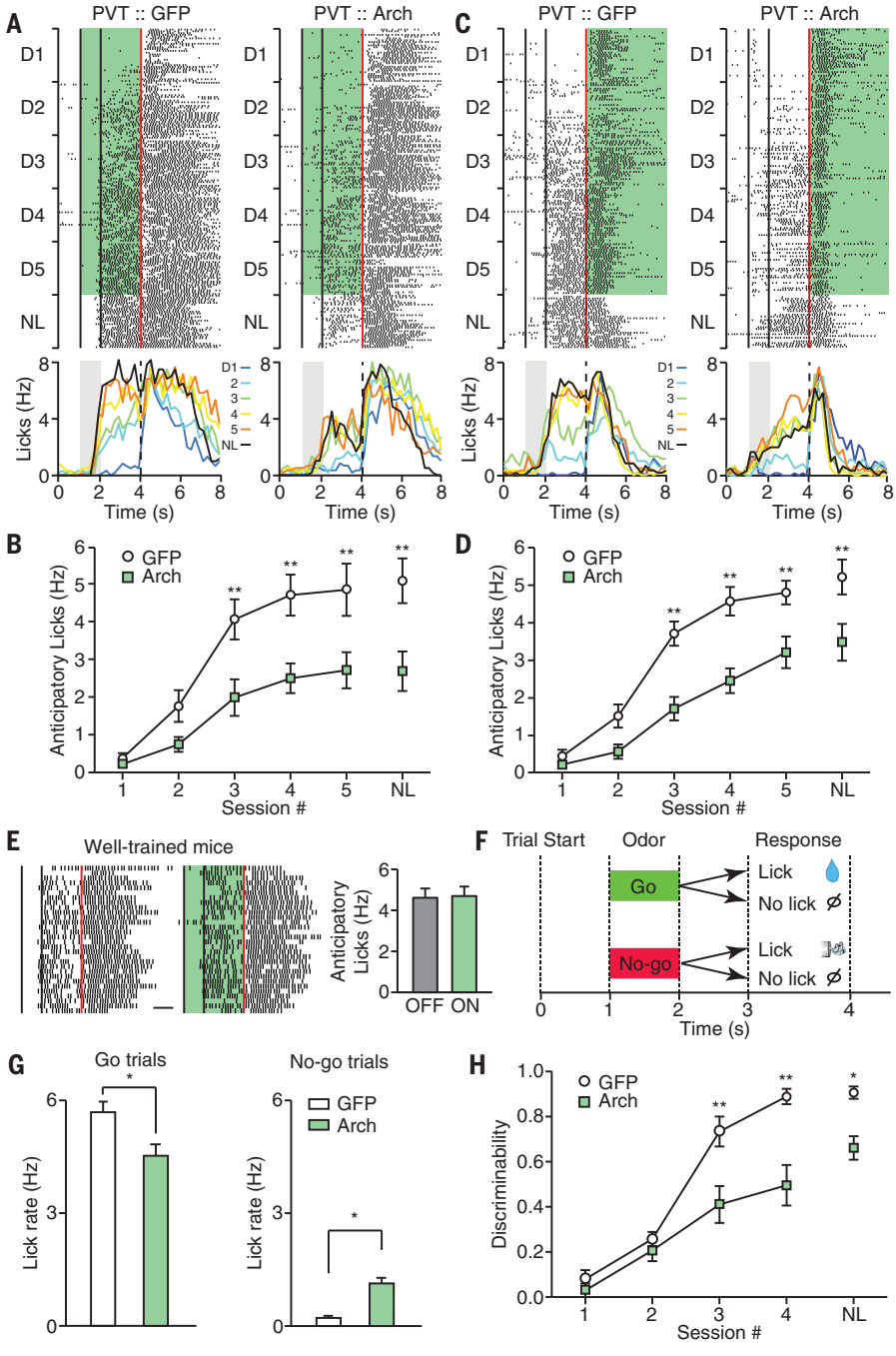


Fig. 3. Photoinhibition of PVT impairs associative learning. (**A** and **C**) (Top) Representative lick raster plots from (left) PVT::GFP and (right) PVT::ArchT mice across five conditioning sessions and the NL test when light stimulation was delivered (A) during CS + delay period or (C) after US delivery. NL, no laser. Back lines indicate the start and end time for odor delivery, respectively. Red line indicates water delivery. Green shade indicates laser stimulation. (Bottom) Representative change of lick rate across five conditioning sessions (D1 to D5, blue, light blue, green, yellow, and orange, respectively) and the NL test (black). (**B** and **D**) Quantification of anticipatory licks of (A) and (C), respectively. (B) $n = 7$ PVT::GFP mice; $n = 7$ PVT::ArchT mice. (D) $n = 6$ PVT::GFP mice; $n = 8$ PVT::ArchT mice. $^{**}P < 0.01$ (two-way ANOVA, post hoc Bonferroni test). (**E**) Representative lick raster plots (left) and histograms (right) from well-trained PVT::ArchT mice with laser off and on. Green, laser on ($n = 6$ mice). Scale bar, 1 s. Wilcoxon signed-rank test, $P = 0.84$. (**F**) Schematics of go/no-go task. (**G**) Anticipatory lick rate of (left) go trials and (right) no-go trials from PVT::GFP ($n = 9$) and PVT::ArchT ($n = 10$) mice on the last day of training. Mann-Whitney U test, $^{*}P < 0.05$. (**H**) Discriminability of go and no-go trials over training. Discriminability was calculated as $(Lick_{go} - Lick_{no\text{-}go})/(Lick_{go} + Lick_{no\text{-}go})$. $^{*}P < 0.05$, $^{**}P < 0.01$ (two-way ANOVA, post-hoc Bonferroni test). Data are means ± SEM.

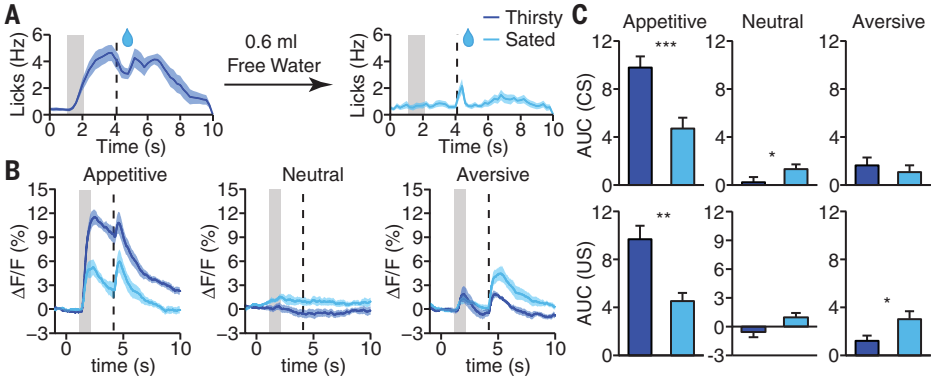


Fig. 4. Effect of homeostatic state on PVT responses. (**A**) Mean lick rate after odor cue in thirsty (left, dark blue, *n* = 7 mice) and sated (right, light blue, *n* = 7 mice) state. (**B** and **C**) Mean photometric traces (B) and quantification (C) of PVT responses in (left) appetitive, (middle) neutral, and (right) aversive test in thirsty (left, dark blue) and sated (right, light blue) state. (C) CS (top) and US (bottom) response. Thirsty, *n* = 7 mice; sated, *n* = 7 mice; **P* < 0.05, ***P* < 0.01, ****P* < 0.001 (Mann-Whitney *U* test). Shade, SEM across mice in (A) and (B). Gray bar indicates 1 s of CS delivery, and vertical dashed line indicates US delivery in (A) and (B). Data are means ± SEM.

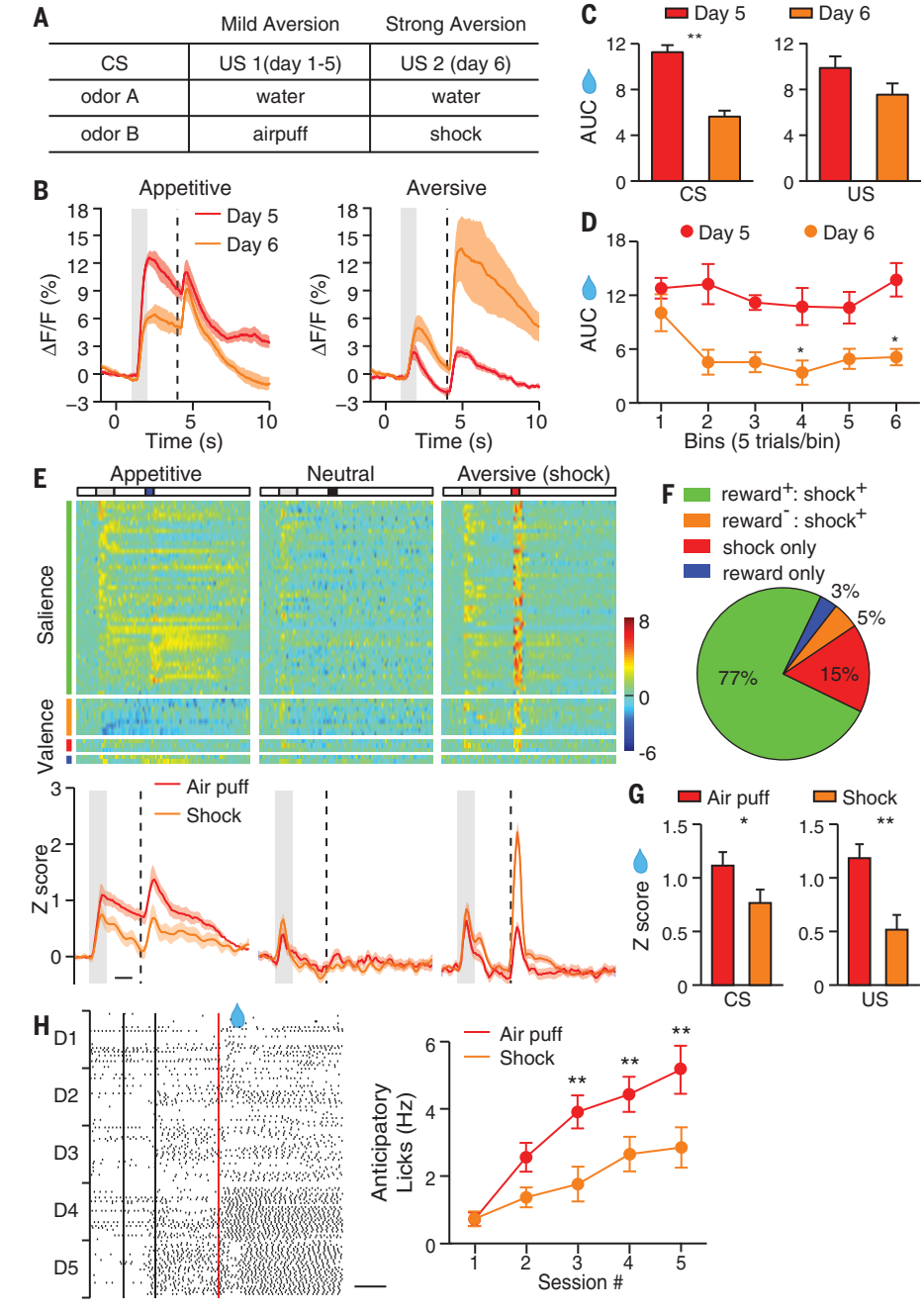


Fig. 5. Context-dependent modulation of salience response in the PVT. (**A**) Behavioral procedure for switching from mild to strong aversive context. (**B**) Mean photometric responses of the PVT to appetitive and aversive test in mild (*n* = 6 mice) versus strong aversive context (*n* = 6 mice). (**C**) Quantification of CS (left) and US (right) response in (B). Mann-Whitney *U* test, ***P* < 0.01 (CS); *P* = 0.15 (US). (**D**) Rapid suppression of PVT response to water-predicting cue after switching from mild to strong aversive context. There was no further reduction observed after 10 trials. (**E**) (Top) *Z* score heat maps for all task-responding neurons identified by means of in vivo single-unit recording of well-trained animals in strong aversive context. Neurons are separated in four subgroups on the basis of their tuning properties and are rank-ordered by their response onset times during reward cue stimulation. Each row in the heat maps represents responses from the same neuron to different stimuli. *n* = 62 neurons from 12 mice. (Bottom) *Z* score quantification of PVT response during Pavlovian tasks. (**F**) Pie chart shows the tuning of PVT neurons in strong aversive context. (**G**) *Z* score quantification of CS (left) and US (right) response in the PVT during appetitive test in mild (*n* = 85 neurons) versus strong aversive context (*n* = 62 neurons). Mann-Whitney *U* test, **P* < 0.05 (CS); ***P* < 0.01 (US). (**H**) (Left) Raster plots illustrate licking behavior across five reward conditioning sessions in strong aversive context. (Right) Quantification of anticipatory licks during reward conditioning in mild (red, *n* = 7 mice) versus strong aversive context (orange, *n* = 7 mice). ***P* < 0.01 (two-way ANOVA, post-hoc Bonferroni test). Red, mild aversive condition; orange, strong aversive condition, in (E), (G), and (H). Shade, SEM across mice in (B) and (E). Gray bar indicates 1 s of CS delivery, and vertical dashed line indicates US delivery in (B) and (E). Scale bar, 1 s in (E) and (H). Data are means ± SEM.

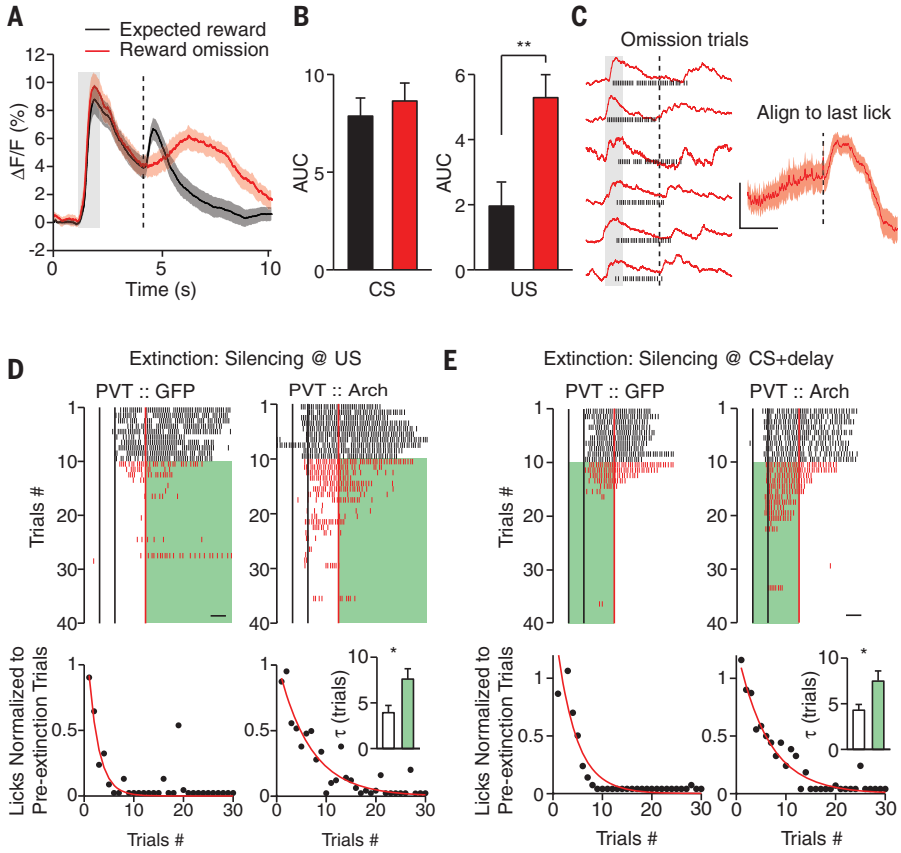


Fig. 6. Reward-omission response in the PVT. (**A** and **B**) Mean photometric traces (A) and histogram (B) illustrating delayed but long-lasting PVT responses to reward omission. Expected reward (black, n = 10 mice); reward omission (red, n = 10 mice), Wilcoxon signed-rank test, P = 0.19 (CS); $^{**}P < 0.01$ (US). (**C**) (Left) Representative traces of individual omission response (red) superimposed with lick raster plots (black). (Right) Mean photometric traces (n = 10 mice) after aligning to the last lick in omission trials. Shown is the rapid increase of calcium signals after licking stops. Scale bar, 2% $\Delta F/F$, 1 s. Gray bar indicates CS delivery, and vertical dashed line indicates US delivery in (A) and (C). (**D** and **E**) (Top) Representative lick raster plots from (left) PVT::GFP and (right) PVT::ArchT mice with laser stimulation during (D) reward-omission period or (E) CS + delay period of extinction trials. Back lines indicate the start and end time for odor delivery, respectively. Red line indicates water delivery. Scale bar, 1 s. The mice received water reward in first 10 trials (black), then water delivery stopped (red), and optogenetic stimulation was on until the end of the trial (green). (Bottom) Quantification of anticipatory licks in 30 extinction trials. Licks (black dot) are normalized to averaged licks during the first 10 trials. Red line indicates the exponential fit of licks. (D) (Inset) Histogram shows the mean time constants (τ) of extinction from PVT::GFP (white, n = 6) and PVT::ArchT (green, n = 10) mice. (E) (Inset) Histogram shows the mean time constants (τ) of extinction from PVT::GFP (white, n = 9) and PVT::ArchT (green, n = 10) mice. Mann-Whitney U test, $^{*}P < 0.05$. Shade, SEM across mice in (A) and (C). Data are means ± SEM.

the PVT. These results demonstrate that the PVT does not encode PE (*6*, *30*, *31*). Moreover, inhibition of PVT activity impairs associative learning of appetitive and aversive outcomes as well as extinction of an established reward association. Together, our results highlight the importance of stimulus salience in driving learning. Silencing PVT activity affects learning but not expression of conditioned behavior, indicating that the function of the PVT is different from other thalamic nuclei such as the mediodorsal thalamus or the thalamus that connects with the anterior lateral motor cortex because silencing these regions disrupts ongoing task performance (*12*–*15*). In well-trained mice, silencing PVT CS response has no effect on licking when water is available but slowed extinction when water is not available, which suggests that CS responses in well-trained mice are for monitoring potential changes of salience. The critical next step is to determine how salience information in the PVT is communicated to the rest of the brain. Axons of PVT neurons show extensive collateralization; therefore, the PVT could simultaneously broadcast salience signals to multiple downstream targets to coordinate their activities (fig. S9) (*38*). PVT terminals in the nucleus accumbens (NAc) directly interact with dopaminergic fibers from the ventral tegmental area and evoke dopamine efflux, suggesting a direct interaction of salience and reward PE signals in the NAc (*39*). The impact of these interactions on associative learning needs to be investigated further.

REFERENCES AND NOTES

1. E. I. Knudsen, *Annu. Rev. Neurosci.* **30**, 57–78 (2007).
2. J. H. Fecteau, D. P. Munoz, *Trends Cogn. Sci.* **10**, 382–390 (2006).
3. V. Menon, *Salience Network: Brain Mapping: An Encyclopedic Reference* (Elsevier, 2015).
4. L. Q. Uddin, *Nat. Rev. Neurosci.* **16**, 55–61 (2015).
5. A. Ghazizadeh, W. Griggs, O. Hikosaka, *Front. Neurosci.* **10**, 378 (2016).
6. W. Schultz, *J. Neurophysiol.* **80**, 1–27 (1998).
7. G. R. Esber, M. Haselgrove, *Proc. Biol. Sci.* **278**, 2553–2561 (2011).
8. G. Hall, G. Rodríguez, *J. Exp. Psychol. Anim. Learn. Cogn.* **43**, 48–61 (2017).
9. J. P. Gottlieb, M. Kusunoki, M. E. Goldberg, *Nature* **391**, 481–484 (1998).
10. L. J. Toth, J. A. Assad, *Nature* **415**, 165–168 (2002).
11. R. Ptak, *Neuroscientist* **18**, 502–515 (2012).
12. L. Acsády, *Nat. Neurosci.* **20**, 901–902 (2017).
13. Z. V. Guo *et al.*, *Nature* **545**, 181–186 (2017).
14. L. I. Schmitt *et al.*, *Nature* **545**, 219–223 (2017).
15. S. S. Bolkan *et al.*, *Nat. Neurosci.* **20**, 987–996 (2017).
16. E. Z. Millan, Z. Ong, G. P. McNally, *Prog. Brain Res.* **235**, 113–137 (2017).
17. J. M. Otis *et al.*, *Nature* **543**, 103–107 (2017).
18. F. H. Do-Monte, A. Minier-Toribio, K. Quiñones-Laracuente, E. M. Medina-Colón, G. J. Quirk, *Neuron* **94**, 388–400.e4 (2017).
19. E. A. Choi, G. P. McNally, *J. Neurosci.* **37**, 3018–3029 (2017).
20. Y. Zhu, C. F. Wienecke, G. Nachtrab, X. Chen, *Nature* **530**, 219–222 (2016).
21. M. A. Penzo *et al.*, *Nature* **519**, 455–459 (2015).
22. G. J. Kirouac, *Neurosci. Biobehav. Rev.* **56**, 315–329 (2015).
23. F. H. Do-Monte, K. Quiñones-Laracuente, G. J. Quirk, *Nature* **519**, 460–463 (2015).
24. D. T. Hsu, G. J. Kirouac, J. K. Zubieta, S. Bhatnagar, *Front. Behav. Neurosci.* **8**, 73 (2014).
25. W. E. Allen *et al.*, *Science* **357**, 1149–1155 (2017).
26. J. N. Betley, Z. F. Cao, K. D. Ritola, S. M. Sternson, *Cell* **155**, 1337–1350 (2013).
27. T. W. Chen *et al.*, *Nature* **499**, 295–300 (2013).
28. G. Cui *et al.*, *Nature* **494**, 238–242 (2013).
29. L. A. Gunaydin *et al.*, *Cell* **157**, 1535–1551 (2014).
30. W. Menegas, B. M. Babayan, N. Uchida, M. Watabe-Uchida, *eLife* **6**, e21886 (2017).
31. J. Y. Cohen, S. Haesler, L. Vong, B. B. Lowell, N. Uchida, *Nature* **482**, 85–88 (2012).
32. S. C. Lin, M. A. Nicolelis, *Neuron* **59**, 138–149 (2008).
33. M. Matsumoto, O. Hikosaka, *Nature* **459**, 837–841 (2009).
34. J. Tian *et al.*, *Neuron* **91**, 1374–1389 (2016).
35. B. Y. Chow *et al.*, *Nature* **463**, 98–102 (2010).
36. M. Ogawa *et al.*, *Neuron* **77**, 251–258 (2013).
37. M. J. Wagner, T. H. Kim, J. Savall, M. J. Schnitzer, L. Luo, *Nature* **544**, 96–100 (2017).
38. X. Dong, S. Li, G. J. Kirouac, *Brain Struct. Funct.* **222**, 3927–3943 (2017).
39. M. P. Parsons, S. Li, G. J. Kirouac, *J. Comp. Neurol.* **500**, 1050–1063 (2007).

ACKNOWLEDGMENTS

We thank E. I. Knudsen, J. R. Huguenard, and members of the Chen laboratory for helpful comments on the manuscript. We thank E. L. Adams for iDISCO experiments, C. Ran for tail-flick test, and H. S. Yao for head-plate. **Funding:** This work was supported by grants from The Whitehall Foundation, Ajinomoto Innovation Alliance Program, Terman Fellowship, Friminich Scholarship, the Stanford Neuroscience Institute's Neurochoice Initiative, and

start-up funding from Stanford University (to X.C.). Y.Z. is partially supported by a National Alliance for Research on Schizophrenia and Depression Young Investigator Grant from the Brain and Behavior Research Foundation, National Natural Science Foundation of China (grant 31741058), and The Thousand Youth Talents Plan. G.N. is supported by a training grant from the National Institute on Drug Abuse (5T32DA035165-02). W.E.A. is supported by a Fannie and John Hertz Foundation Fellowship and a NSF Graduate Research Fellowship. L.L. is an investigator of the Howard Hughes Medical Institute. **Author contributions:** X.C. conceived the study. Y.Z., W.E.A., L.L., and X.C. designed the experiments. Y.Z and P.C.K. conducted experiments. Y.Z. and X.C. analyzed data. Y.Z., G.N., and X.C. wrote the paper with inputs from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6413/423/suppl/DC1
Materials and Methods
Figs. S1 to S9
Table S1

18 January 2018; accepted 28 August 2018
10.1126/science.aat0481

REPORT

NEUROSCIENCE

The paraventricular thalamus is a critical thalamic area for wakefulness

Shuancheng Ren[1]*, Yaling Wang[1]*, Faguo Yue[1,2]*, Xiaofang Cheng[1], Ruozhi Dang[1], Qicheng Qiao[1], Xueqi Sun[1], Xin Li[1], Qian Jiang[2], Jiwei Yao[3], Han Qin[3], Guanzhong Wang[1], Xiang Liao[3], Dong Gao[2], Jianxia Xia[1], Jun Zhang[1], Bo Hu[1], Junan Yan[3], Yanjiang Wang[4], Min Xu[5], Yunyun Han[6], Xiangdong Tang[7], Xiaowei Chen[3]†, Chao He[1]†, Zhian Hu[1]†

Clinical observations indicate that the paramedian region of the thalamus is a critical node for controlling wakefulness. However, the specific nucleus and neural circuitry for this function remain unknown. Using in vivo fiber photometry or multichannel electrophysiological recordings in mice, we found that glutamatergic neurons of the paraventricular thalamus (PVT) exhibited high activities during wakefulness. Suppression of PVT neuronal activity caused a reduction in wakefulness, whereas activation of PVT neurons induced a transition from sleep to wakefulness and an acceleration of emergence from general anesthesia. Moreover, our findings indicate that the PVT–nucleus accumbens projections and hypocretin neurons in the lateral hypothalamus to PVT glutamatergic neurons' projections are the effector pathways for wakefulness control. These results demonstrate that the PVT is a key wakefulness-controlling nucleus in the thalamus.

Patients with occlusion of the paramedian thalamic artery, which results in localized injury to the paramedian region of the thalamus, show disturbances of consciousness ranging from hypersomnolence to sleep-like coma when injuries are bilateral (*1*–*4*), a clinical feature not observed in other thalamic injuries (*2*). The homologous area of the primate paramedian thalamus in rodents consists of a large number of nuclei, including the paraventricular thalamus (PVT), nucleus reuniens, mediodorsal nucleus, and interanteromedial thalamic nucleus (*5*, *6*). These nuclei have distinct input and output connections (*6*–*8*) and participate in various brain functions (*5*, *9*, *10*). However, the specific nucleus and circuitry controlling wakefulness have not yet been identified.

We began by visualizing an unbiased map of c-fos expression in the paramedian thalamus after a period of wakefulness or sleep in mice. We observed a higher level of c-fos expression in the PVT than in the other nuclei of the paramedian thalamus at zeitgeber time 18 (ZT 18; 24:00) and after extended wakefulness (fig. S1). To further examine the in vivo dynamics of PVT neurons during the sleep-wake cycle, we injected adeno-associated virus (AAV) expressing the genetically encoded Ca^{2+} sensor GCaMP6f under the control of the CaMKIIα promoter (PVT neurons are primarily glutamatergic neurons) (*11*) into the PVT (fig. S2, A and B). The population Ca^{2+} activity was significantly higher during wakefulness than during sleep. The population Ca^{2+} activity of the PVT began to increase ahead of signs of behavioral arousal (fig. S2, C to H). We next performed in vivo multichannel electrophysiological recordings to monitor the spike firing of individual PVT neurons in freely behaving mice (Fig. 1A). The vast majority (20/22) of PVT neurons exhibited a higher firing rate during wakefulness than during sleep (Fig. 1, B and C). The PVT firing rate gradually decreased before sleep onset and increased during transitions from sleep to wakefulness. At the onset of behavioral arousal from non–rapid eye movement (NREM) sleep, the mean firing rate reached 7.1 Hz (Fig. 1, D to F). In a considerable fraction of neurons (25%), the firing rate exceeded 10 Hz (fig. S3). The increase of PVT neuronal firing occurred briefly before both cortical activation (1.0 ± 0.3 s) and behavioral arousal (1.4 ± 0.3 s).

We next determined the necessity of PVT activity for wakefulness by inhibiting PVT glutamatergic neurons chemogenetically with AAV encoding engineered G_i-coupled hM4D receptor (AAV-CaMKIIα-hM4D-mCherry) (Fig. 1G). Whole-cell recordings of PVT neurons from acute brain slices confirmed that clozapine-*N*-oxide (CNO, 5 μM) potently inhibited hM4D-expressing PVT neurons (fig. S4). At the beginning of the dark phase (ZT 12; 18:00), CNO injection induced a significant reduction in wakefulness relative to the mCherry and saline controls (Fig. 1, H and I, and fig. S6A), which was primarily due to shortened duration of wakefulness episodes and increased sleep episodes (fig. S5). In addition, spectral analysis of electroencephalography (EEG) showed that CNO injection increased the high delta power (2 to 4 Hz) of wakefulness at hour 2 after CNO injection (figs. S6 and S7). Note that inhibition of the PVT increased the number of micro-arousals (fig. S8, A and B), which might result from a fragmentation of wakefulness. In contrast, CNO injection at the beginning of the light phase (ZT 0; 6:00) did not further reduce wakefulness (fig. S9, A, D, and G). The EEG power spectrum for each state and the number of micro-arousals during the light phase were not affected by CNO injection (figs. S8C and S9).

We next explored the role of the PVT in controlling wakefulness by ablating PVT glutamatergic neurons. We injected a mixture of AAV encoding diphtheria toxin A (AAV-DIO-DTA) and AAV-CaMKIIα-Cre-GFP into the PVT. PVT glutamatergic neurons were selectively ablated after 4 weeks of virus injection (fig. S10A). Animals with chronic PVT lesion showed a decrease of wakefulness during the dark phase, with a reduction in duration of wakefulness episodes and an increase in NREM sleep episodes and micro-arousals, whereas lesion of the PVT did not affect wakefulness during the light phase (fig. S10, B to G). DTA lesion increased the EEG high delta power (2 to 4 Hz) of wakefulness during the dark phase. This lesion also caused an increase in the low theta power (4 to 7 Hz) (fig. S10, H to L), which might reflect a compensatory mechanism during chronic ablation of neurons induced by DTA. In addition, we also injected ibotenic acid to rapidly ablate PVT neurons within several days (Fig. 1J and fig. S11A). Acute PVT lesion reduced wakefulness during the dark phase but did not affect wakefulness during the light phase (Fig. 1K and fig. S11, B and C). Lesion of the PVT caused fragmentation of wakefulness, which was indicated by shortened duration of wakefulness episodes and by increased sleep bouts and micro-arousals during the dark phase. In contrast to chronic lesion, ibotenic acid lesion increased the EEG high delta power (2 to 4 Hz) but not the theta power, indicating a more typical damage to the wakefulness state (Fig. 1L, fig. S11, D to J, and movie S1).

We then used optogenetics to examine the causal role of the PVT in wakefulness control. We injected AAV expressing channelrhodopsin 2 (AAV-CaMKIIα-ChR2-mCherry) into the PVT (Fig. 2A). Functional expression of ChR2 was verified by in vitro electrophysiology (Fig. 2B). We applied optical stimulation (lasting 20 s) after the onset of stable NREM or REM sleep during the light phase. Optical stimulation of PVT glutamatergic neurons during NREM sleep

[1]Department of Physiology, Collaborative Innovation Center for Brain Science, Third Military Medical University, Chongqing 400038, China. [2]Department of Sleep and Psychology, Daping Hospital, Third Military Medical University, Chongqing 400042, China. [3]Brain Research Center, Third Military Medical University, Chongqing 400038, China. [4]Department of Neurology, Daping Hospital, Third Military Medical University, Chongqing 400042, China. [5]Institute of Neuroscience, State Key Laboratory of Neuroscience, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China. [6]Department of Neurobiology, School of Basic Medicine and Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China. [7]Sleep Medicine Center, Laboratory of Anaesthesia and Critical Care Medicine, Translational Neuroscience Center, West China Hospital, Sichuan University, Chengdu 610041, China.
*These authors contributed equally to this work.
†Corresponding author. Email: xiaowei_chen@tmmu.edu.cn (X.C.); hechaochongqing@163.com (C.H.); zhianhu@aliyun.com (Z.H.)

reliably induced transitions to wakefulness in a frequency-dependent manner (Fig. 2, C and D). Optical stimulation of the PVT at a shorter duration (lasting 5 s) was still sufficient to induce such transitions (fig. S12). Optical stimulation also induced transitions to wakefulness from REM sleep (fig. S13, A and B). Optical stimulation significantly increased the probability of wakefulness, along with a complementary decrease of both NREM and REM sleep probability, relative to the mCherry control (Fig. 2E and fig. S13, C and D). We also delivered prolonged optical stimulation to test the ability of PVT neurons in maintaining wakefulness (fig. S14A). Such optical stimulation resulted in an

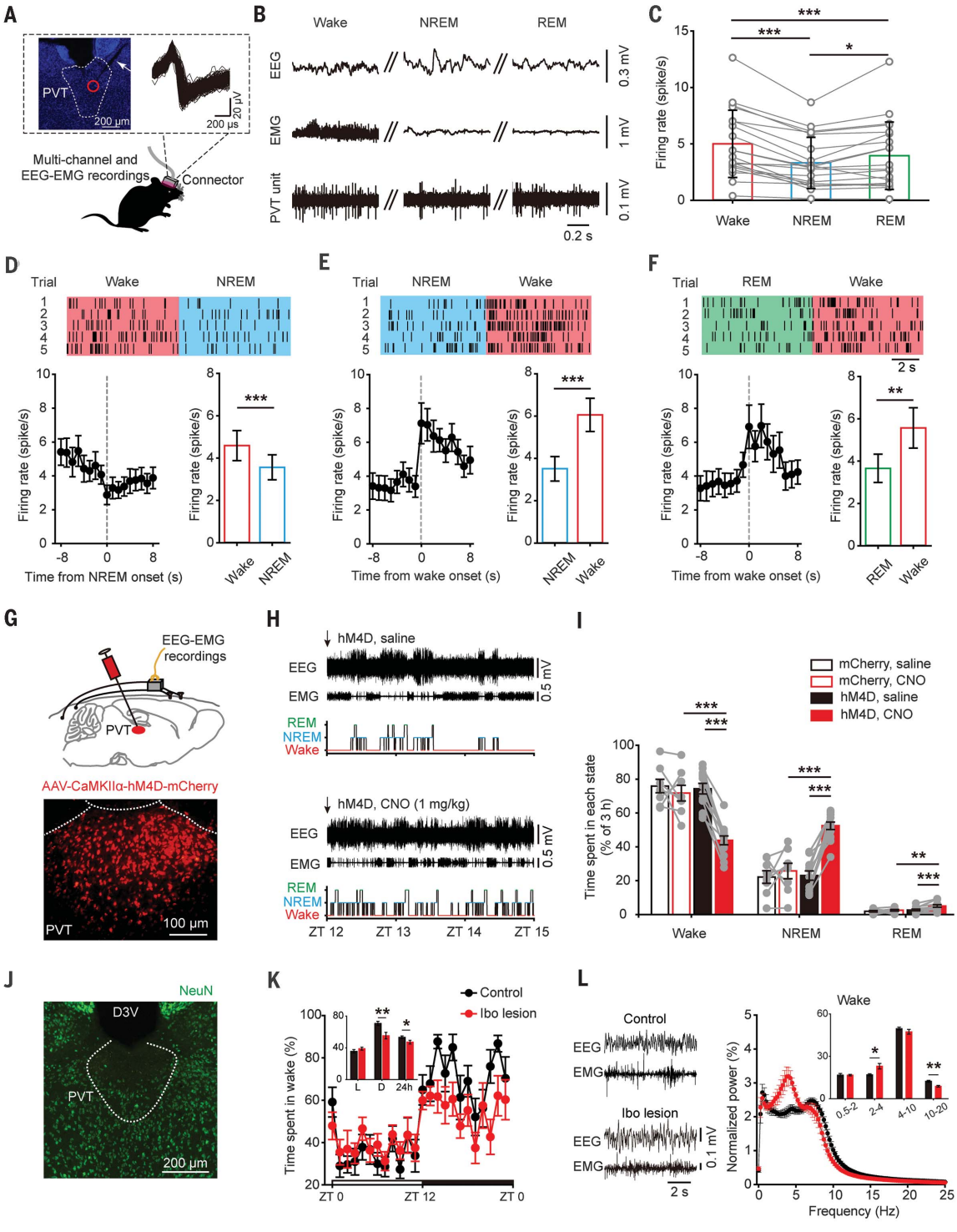


Fig. 1. PVT glutamatergic neurons are required for the control of wakefulness. (**A**) Schematic configuration of in vivo multichannel electrophysiological recordings. Left inset: A brain slice from a mouse with electrodes implanted in the PVT. White arrow indicates the electrode track. Red circle indicates the recording position. Right inset: Waveforms from a recorded PVT neuron. (**B**) EEG, electromyography (EMG), and PVT unit recording traces during wakefulness, NREM sleep, and REM sleep. (**C**) Average firing rate of PVT neurons in each state. (**D** to **F**) Firing rate of PVT neurons during state transitions: wake-to-NREM (D), NREM-to-wake (E), and REM-to-wake (F). Top: Example rastergrams of a PVT neuron during five trials of different state transitions. Bottom left: Average firing rate during the state transition period. Bottom right: Average firing rate of 8 s before and after state transitions. (**G**) Top: Schematic of virus injection and EEG-EMG recordings. Bottom: Image showing the expression of hM4D-mCherry in PVT neurons. (**H**) EEG-EMG traces and hypnograms during 3 hours after saline or CNO (1 mg/kg) injection in an hM4D-mCherry mouse. (**I**) Percentage of time spent in each state during 3 hours after injection. (**J**) Image showing NeuN (neuron-specific nuclear protein) staining from a mouse with PVT lesion using ibotenic acid (Ibo). (**K**) Hourly percentage of time spent in wakefulness across the 24-hour sleep-wake cycle. L, light phase; D, dark phase. (**L**) Left: Raw EEG-EMG traces of wakefulness. Right: Normalized EEG power density of wakefulness during the dark phase. Inset in (L) is a quantitative analysis of the power in different frequency bands. *P < 0.05, **P < 0.01, ***P < 0.001. Error bars denote SEM. See table S1 for further details of statistical data analysis.

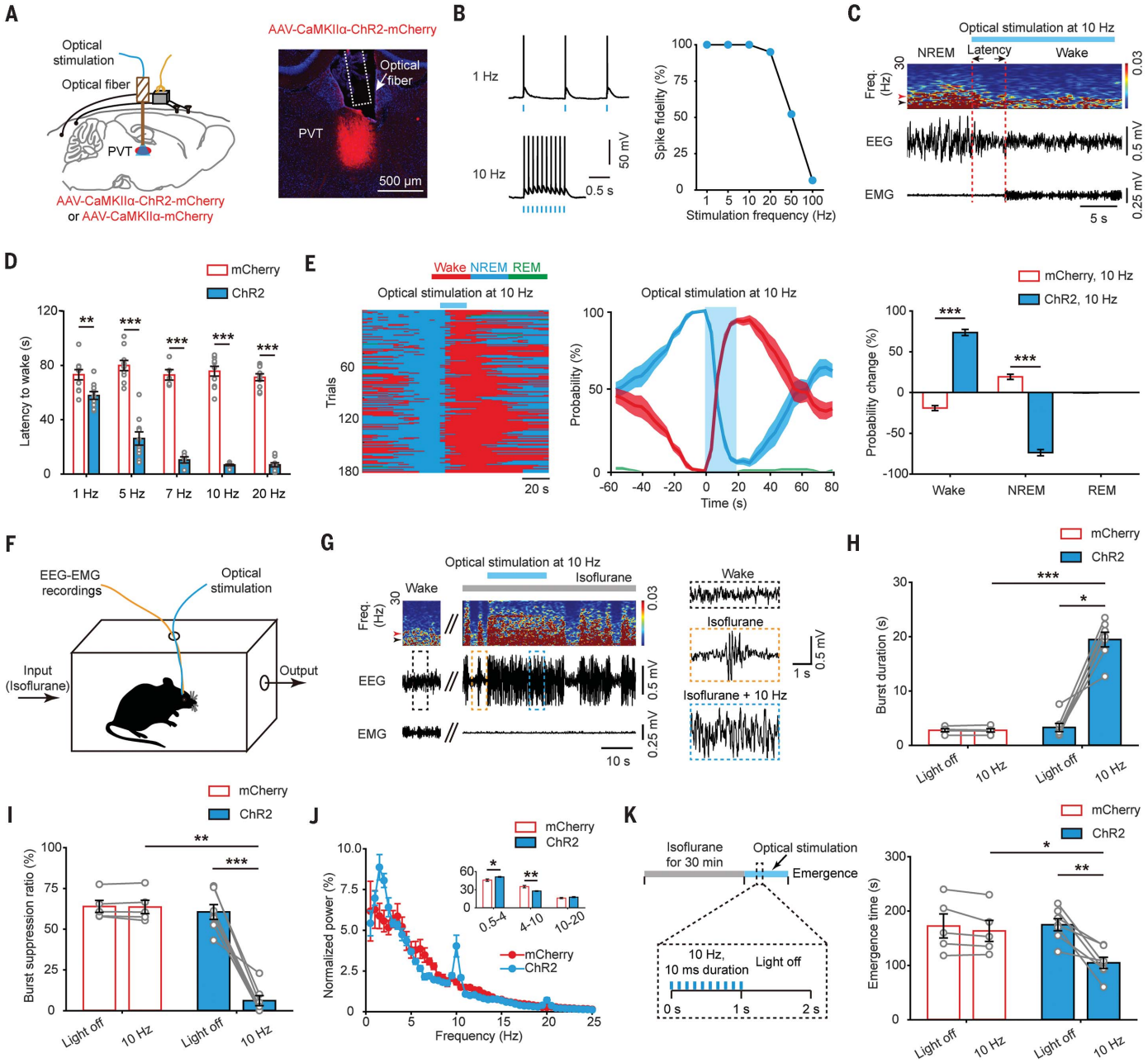


Fig. 2. Optogenetic activation of PVT glutamatergic neurons induces wakefulness from sleep and general anesthesia. (**A**) Left: Schematic of optogenetic manipulation of PVT glutamatergic neurons and EEG-EMG recordings. Right: ChR2-mCherry expression and location of optical fiber in the PVT. (**B**) Example traces (left) and fidelity of action potential firing (right) of ChR2-expressing PVT neurons evoked by 473-nm light stimulation with different frequencies. (**C**) Representative EEG power spectrum and EEG-EMG traces around 10-Hz stimulation delivered during NREM sleep. Arrowheads indicate 4 Hz and 8 Hz. Color scale indicates the power (mV^2) of raw power spectral density. Fre., frequency. (**D**) Latency to wakefulness from NREM sleep after optical activation at 1 Hz, 5 Hz, 7 Hz, 10 Hz, or 20 Hz. (**E**) Left: Brain states in all trials from ChR2-expressing mice. Center: Probability of each state around optogenetic stimulation delivered during NREM sleep. Right: Probability change of each state 20 s before and during optical stimulation. (**F**) Experimental setup for EEG-EMG recordings with simultaneous optogenetic activation in the presence of continuous isoflurane. (**G**) Left: EEG power spectrum and EEG-EMG traces around 10-Hz stimulation during isoflurane-induced general anesthesia. Right: Enlarged view of EEG activity during wakefulness, isoflurane-induced anesthesia, and 10-Hz optical stimulation in continuous isoflurane conditions. (**H** and **I**) Burst duration (H) and burst suppression ratio (I) during periods of light off and 10-Hz optical stimulation. (**J**) Spectral analysis of EEG activity during optical stimulation in the presence of isoflurane. (**K**) Left: Schematic of sustained optical stimulation protocol after the discontinuation of isoflurane. Right: Emergence time during light off and 10-Hz optical stimulation. *P < 0.05, **P < 0.01, ***P < 0.001. Error bars denote SEM. See table S1 for further details of statistical data analysis.

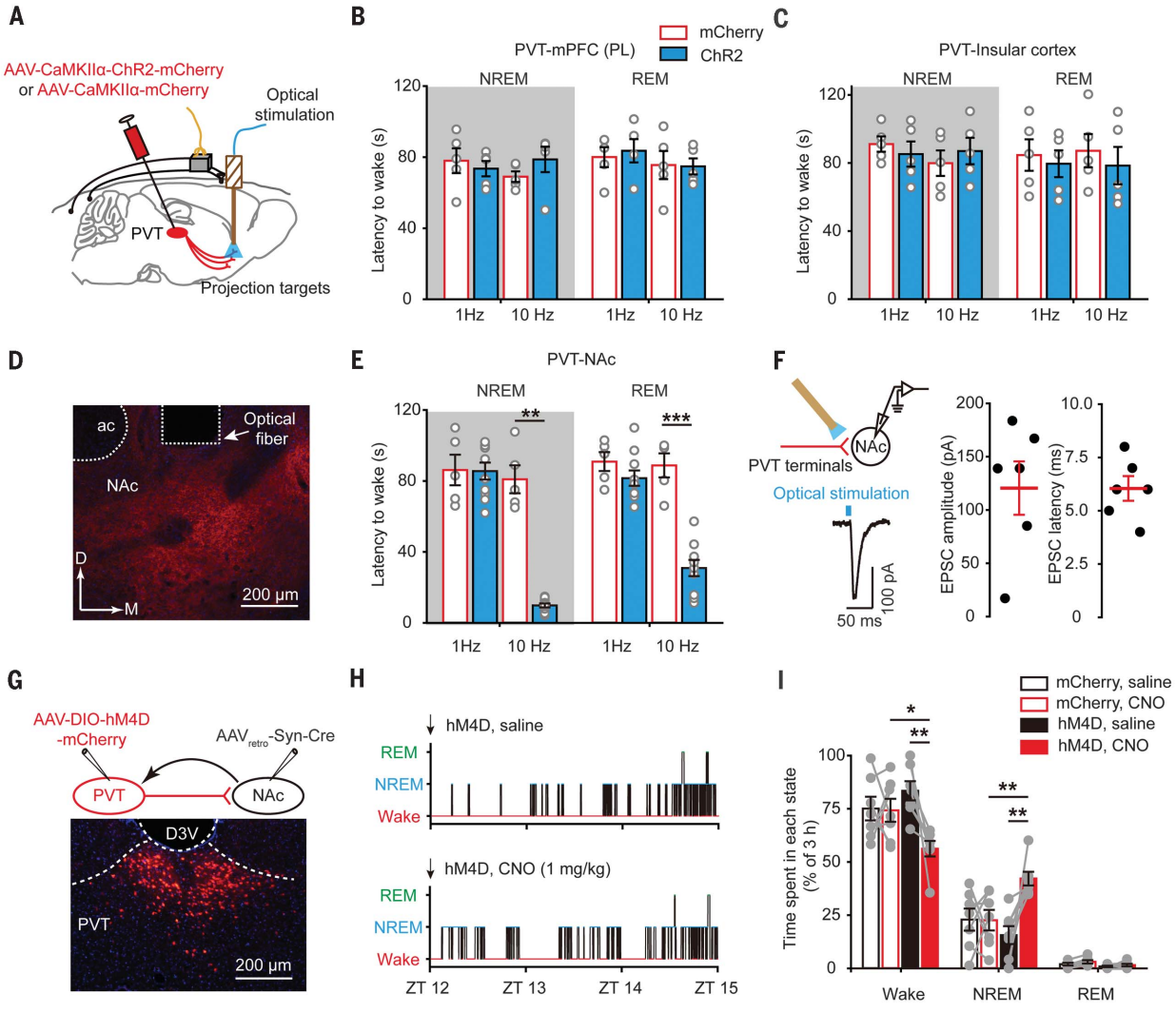


Fig. 3. PVT glutamatergic neurons control wakefulness through the PVT-NAc pathway. (**A**) Schematic of optogenetic stimulation of ChR2-expressing PVT glutamatergic terminals together with EEG-EMG recordings. (**B** and **C**) Latency to wakefulness from NREM and REM sleep after stimulation of PVT projections to the prelimbic cortex (PL) of the mPFC (B) and to the insular cortex (C) at 1 or 10 Hz. (**D**) Distribution of ChR2-expressing PVT glutamatergic terminals and location of optical fiber in the NAc. D, dorsal; M, medial; ac, anterior commissure. (**E**) Latency to wakefulness from NREM and REM sleep after activation of the PVT-NAc pathway at 1 or 10 Hz. (**F**) Top left: Experimental paradigm for in vitro characterization of functional connection of the PVT to NAc. Bottom left: Trace of the single optical stimulation–evoked excitatory postsynaptic currents (EPSCs) in NAc neurons. Right: Amplitude and latency of optical stimulation–evoked EPSCs. (**G**) Top: Experimental design of chemogenetic inhibition of the NAc-projecting PVT neurons. Bottom: Expression of hM4D-mCherry in NAc-projecting PVT neurons. D3V, dorsal third ventricle. (**H**) Hypnograms of an hM4D-mCherry mouse during 3 hours after saline or CNO injection. (**I**) Percentage of time spent in each state during 3 hours after injection. *P < 0.05, **P < 0.01, ***P < 0.001. Error bars denote SEM. See table S1 for further details of statistical data analysis.

overt increase in wakefulness with a decrease of EEG delta power (fig. S14, B to F).

To investigate whether activation of PVT neurons is sufficient to induce wakefulness from an unconscious state—a typical feature of patients with paramedian thalamic stroke (*4*)—we optically activated PVT neurons in mice under general anesthesia induced by isoflurane (Fig. 2F). When a stable EEG burst-suppression mode (a marker of anesthetic depth) (*12*) was observed, 10-Hz stimulation (lasting 20 s) caused an immediate increase in total EEG burst activity. The increase in burst activity was accompanied by a prolonged burst duration, a decreased burst suppression ratio, and an altered EEG spectrum (Fig. 2, G to J). Interestingly, a peak at approximately 10 Hz appeared in the EEG spectrum with optical stimulation (Fig. 2J), which might reflect that optogenetic stimulation at 10 Hz causes prominent 10-Hz oscillations in cortex. Additionally, sustained activation of PVT neurons significantly accelerated the emergence from isoflurane-induced unconsciousness, which was not observed in mCherry mice (Fig. 2K).

Recent studies have reported that thalamic nuclei directly regulate the activities of cortical neurons (*13*, *14*). The PVT also sends brainwide projections, including direct projections to the cortex (*15*). Thus, we next searched for the PVT downstream pathways mediating the wakefulness-maintaining effects. ChR2-mCherry expression indicated that PVT glutamatergic neurons send projections to multiple cortical regions, including the medial prefrontal cortex (mPFC) and insular cortex (fig. S15, A to D). The PVT sends dense projections to the nucleus accumbens (NAc) (Fig. 3D), which participates in wakefulness control (*16*). Bilateral optical stimulation of PVT projections to different layers or subregions of the prelimbic cortex of the mPFC failed to induce rapid transitions from NREM sleep to wakefulness. Optical stimulation of PVT axonal terminals in other cortical regions, including the cingulate cortex and infralimbic cortex of the mPFC or the anterior and posterior

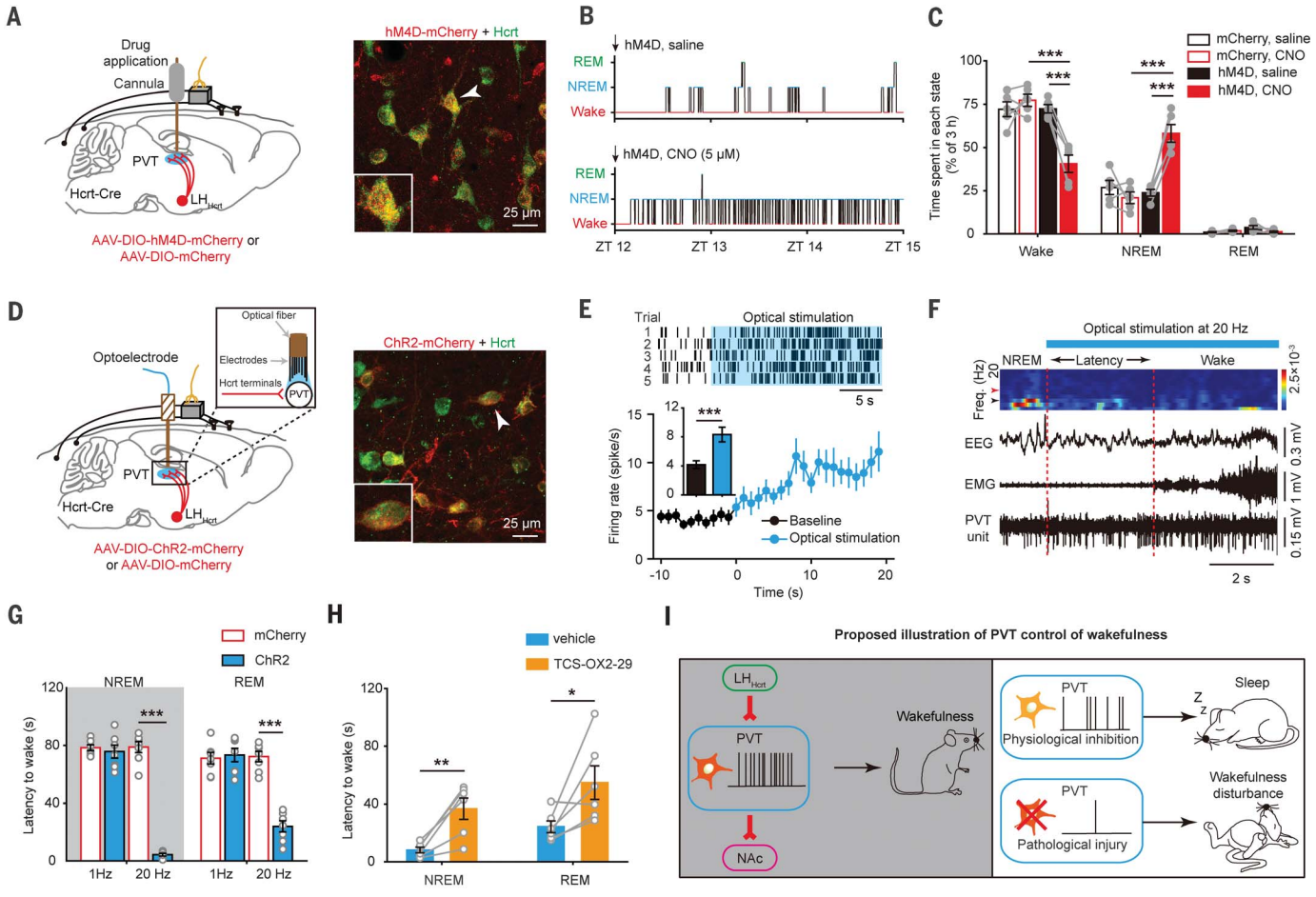


Fig. 4. The wakefulness-controlling function of the PVT is regulated by Hcrt neurons. (**A**) Left: Schematic of chemogenetic inhibition of LH_{Hcrt}-PVT pathway and EEG-EMG recordings. Right: Selective expression of hM4D-mCherry in Hcrt neurons in the LH. Arrowhead denotes neuron shown in close-up (inset). (**B**) Hypnograms of an hM4D-mCherry mouse during 3 hours after saline or CNO (5 μM) local infusion. (**C**) Percentage of time spent in each state during 3 hours after injection. (**D**) Left: Schematic of optogenetic stimulation of LH_{Hcrt}-PVT pathway together with PVT unit and EEG-EMG recordings. Right: ChR2-mCherry expression in Hcrt neurons in the LH. Arrowhead denotes neuron shown in close-up (inset). (**E**) Top: Rastergrams showing the firing activity of a PVT neuron before and during optical stimulation. Bottom: Optical stimulation of Hcrt neurons' terminals significantly increased the average firing rate of PVT neurons. (**F**) EEG-EMG traces and PVT unit firing around 20-Hz optical stimulation in a freely moving mouse. (**G**) Latency to wakefulness from NREM or REM sleep after optical stimulation at 1 or 20 Hz. (**H**) The Hcrt receptor antagonist TCS-OX2-29 attenuated the wakefulness-inducing effects of optical stimulation of Hcrt neuron's terminals in the PVT. (**I**) Model of the PVT control of wakefulness. Left: Increased levels of Hcrt may activate Hcrt inputs to the PVT, which could excite PVT neurons projecting to the NAc, thereby activating NAc neurons to control wakefulness. Top right: Decreased activity of PVT neurons leads to sleep. Bottom right: The impairment of PVT neurons seen in neurological diseases may be associated with disturbances of wakefulness. *P < 0.05, **P < 0.01, ***P < 0.001. Error bars denote SEM. See table S1 for further details of statistical data analysis.

part of the insular cortex, had no obvious effects on sleep-wake transition (Fig. 3, A to C, and fig. S15, E and F), although the paramedian thalamus can excite these cortices via glutamatergic connections (*17*). Moreover, chemogenetic inhibition of the mPFC-projecting PVT neurons also did not affect the time spent in wakefulness (fig. S16). Together, our results do not sufficiently support a crucial role of the PVT-cortex pathway in controlling wakefulness. However, optical stimulation of the PVT-to-NAc projections reliably elicited transitions to wakefulness from both NREM and REM sleep (Fig. 3E). In vitro ChR2-assisted circuit mapping (*18*) confirmed this functional monosynaptic connection (Fig. 3F), indicating that a direct PVT-NAc pathway may mediate the observed transitions. In addition, chemogenetic inhibition of NAc-projecting PVT neurons significantly reduced wakefulness (Fig. 3, G to I). The importance of this pathway in wakefulness control was further confirmed by the compromised wakefulness-inducing effects of the PVT after ablation of NAc neurons using ibotenic acid (fig. S17). Manipulation of the PVT-NAc pathway efficiently reproduced the wakefulness-controlling effects of the PVT (fig. S18).

Because the PVT receives inputs from the brainstem and lateral hypothalamus (LH) (*19*, *20*), we next searched for the upstream pathway that modulates its wakefulness-maintaining function. We used Cre-dependent rabies virus–mediated monosynaptic retrograde tracing in vGlut2-Cre mice and found that PVT glutamatergic neurons received direct inputs from LH hypocretin (Hcrt) neurons (fig. S19), which are involved in wakefulness control (*21*). We specifically suppressed inputs from Hcrt neurons by expressing AAV-DIO-hM4D-mCherry in the LH of Hcrt-Cre mice and locally applying CNO (5 μM) in the PVT (Fig. 4A). CNO infusion significantly decreased wakefulness, accompanied by fragmentation of wakefulness (Fig. 4, B and C, and fig. S20, A to D). Moreover, inhibition of this pathway shortened NREM sleep latency and increased the number of NREM sleep episodes (fig. S20, E and F).

We then examined the sufficiency of these inputs for wakefulness control. Optical stimulation (20 Hz) of Hcrt neurons' terminals during NREM sleep significantly increased the firing rate of PVT neurons (Fig. 4, D and E). At the behavioral level, apparent transitions from sleep

to wakefulness occurred after stimulation of the LH_{Hcrt}-PVT pathway (Fig. 4, F and G, and fig. S21). Optogenetic stimulation of Hcrt neurons' terminals in the PVT could still induce transitions from sleep to wakefulness after blockade of the potential antidromic action potentials by injection of muscimol into the LH (fig. S22). Such wakefulness-controlling effects were partially attenuated by the Hcrt receptor 2 antagonist TCS-OX2-29 (Fig. 4H). To further confirm the importance of the LH_{Hcrt}-PVT pathway in controlling wakefulness, we injected viruses expressing hM3D into the LH and hM4D into the PVT, respectively (fig. S23, A and B). Stimulation of Hcrt neurons increased the amount of wakefulness, whether or not the PVT was chemogenetically inhibited (fig. S23C). These unexpected results might be because the activity of the PVT was not completely suppressed under the condition of activating Hcrt neurons' terminals. Alternatively, after ablating PVT neurons using caspase-3, the latency to wakefulness with optogenetic stimulation of Hcrt neurons increased, confirming an important role of the LH_{Hcrt}-PVT pathway in wakefulness control (fig. S23, D to G).

Recently, the PVT in the paramedian thalamus has been extensively studied because of its involvement in multiple behaviors, including fear conditioning (*22*, *23*), drug addiction (*24*, *25*), and feeding (*26*), all of which require elevated wakefulness. Our results demonstrate that the PVT is both necessary and sufficient for the control of wakefulness, and further reveal the important role of the LH_{Hcrt}-PVT-NAc pathway in the control of wakefulness (Fig. 4I). These results provide experimental evidence supporting the PVT as a critical thalamic node in the wakefulness-controlling neural network.

REFERENCES AND NOTES

1. G. Percheron, *Z. Neurol.* **205**, 1–13 (1973).
2. J. D. Schmahmann, *Stroke* **34**, 2264–2278 (2003).
3. A. Honig *et al.*, *J. Clin. Neurosci.* **34**, 81–85 (2016).
4. D. M. Hermann *et al.*, *Stroke* **39**, 62–68 (2008).
5. R. P. Vertes, S. B. Linley, W. B. Hoover, *Neurosci. Biobehav. Rev.* **54**, 89–107 (2015).
6. H. J. Groenewegen, H. W. Berendse, *Trends Neurosci.* **17**, 52–57 (1994).
7. H. Jasper, *Electroencephalogr. Clin. Neurophysiol.* **1**, 405–419 (1949).
8. K. E. Krout, R. E. Belzer, A. D. Loewy, *J. Comp. Neurol.* **448**, 53–101 (2002).
9. W. Xu, T. C. Südhof, *Science* **339**, 1290–1295 (2013).
10. Y. D. Van der Werf, M. P. Witter, H. J. Groenewegen, *Brain Res. Rev.* **39**, 107–140 (2002).
11. C. Frassoni, R. Spreafico, M. Bentivoglio, *Exp. Brain Res.* **115**, 95–104 (1997).
12. P. C. Vijn, J. R. Sneyd, *Br. J. Anaesth.* **81**, 415–421 (1998).
13. S. Honjoh *et al.*, *Nat. Commun.* **9**, 2100 (2018).
14. T. C. Gent, M. Bandarabadi, C. G. Herrera, A. R. Adamantidis, *Nat. Neurosci.* **21**, 974–984 (2018).
15. R. P. Vertes, W. B. Hoover, *J. Comp. Neurol.* **508**, 212–237 (2008).
16. Y. J. Luo *et al.*, *Nat. Commun.* **9**, 1576 (2018).
17. P. M. Fogerson, J. R. Huguenard, *Neuron* **92**, 687–704 (2016).
18. L. Petreanu, D. Huber, A. Sobczyk, K. Svoboda, *Nat. Neurosci.* **10**, 663–668 (2007).
19. S. Li, G. J. Kirouac, *Brain Struct. Funct.* **217**, 257–273 (2012).
20. G. J. Kirouac, M. P. Parsons, S. Li, *Brain Res.* **1059**, 179–188 (2005).
21. J. Li, Z. Hu, L. de Lecea, *Br. J. Pharmacol.* **171**, 332–350 (2014).
22. M. A. Penzo *et al.*, *Nature* **519**, 455–459 (2015).
23. F. H. Do-Monte, K. Quiñones-Laracuente, G. J. Quirk, *Nature* **519**, 460–463 (2015).
24. Y. Zhu, C. F. Wienecke, G. Nachtrab, X. Chen, *Nature* **530**, 219–222 (2016).
25. G. J. Kirouac, *Neurosci. Biobehav. Rev.* **56**, 315–329 (2015).
26. X. Zhang, A. N. van den Pol, *Science* **356**, 853–859 (2017).

ACKNOWLEDGMENTS

We thank L. de Lecea for providing the Hcrt-Cre mice, and M. M. Luo and Y. M. Lu for critical comments on the manuscript. **Funding:** Supported by the National Natural Science Foundation of China (31771192 to Z.H., 31500862 to C.H., and 81671106 to X.W.C.) and "973 Program" (2015CB759500 to X.W.C.). **Author contributions:** S.R., C.H., and Z.H. designed the study; S.R., Y.L.W., F.Y., X.F.C., R.D., Q.Q., H.Q., J.W.Y., J.A.Y., G.W., X. Li, Y.J.W., and C.H. executed the experiments and performed statistical analysis; Q.J., X.S., X. Liao, D.G., J.X., J.Z., B.H., X.T., and X.W.C. analyzed the data; S.R., C.H., and Z.H. wrote the paper with the help of Y.J.W., M.X., Y.H., X.T., and X.W.C.; and all authors read and commented on the manuscript. Z.H. supervised all aspects of the project. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data necessary to understand and assess the conclusions of this manuscript are available in the paper and the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6413/429/suppl/DC1
Materials and Methods
Figs. S1 to S25
Table S1
References (*27*–*40*)
Movie S1

8 March 2018; accepted 6 September 2018
10.1126/science.aat2512

ORGANIC CHEMISTRY

Efficient access to unprotected primary amines by iron-catalyzed aminochlorination of alkenes

Luca Legnani[1]*, Gabriele Prina-Cerai[1]*, Tristan Delcaillau[1,2]†, Suzanne Willems[1]†, Bill Morandi[1,2]‡

Primary amines are essential constituents of biologically active molecules and versatile intermediates in the synthesis of drugs and agrochemicals. However, their preparation from easily accessible alkenes remains challenging. Here, we report a general strategy to access primary amines from alkenes through an operationally simple iron-catalyzed aminochlorination reaction. A stable hydroxylamine derivative and benign sodium chloride act as the respective nitrogen and chlorine sources. The reaction proceeds at room temperature under air; tolerates a large scope of aliphatic and conjugated alkenes, including densely functionalized substrates; and provides excellent anti-Markovnikov regioselectivity with respect to the amino group. The reactivity of the 2-chloroalkylamine products, an understudied class of amphoteric molecules, enables facile access to linear or branched aliphatic amines, aziridines, aminonitriles, azido amines, and homoallylic amines.

Aliphatic primary amines are present in a wide range of important bioactive compounds. They are also among the most versatile synthetic intermediates in the atom-economical construction of secondary amines, tertiary amines, and heterocycles (*1*). The synthesis of primary amines traditionally relies on reductive amination (*2*, *3*), dehydrogenative coupling of alcohols (*4*), allylic amination (*5*, *6*), or azide or nitrile reduction. These methods rely on the preinstallation of a polar group in the starting material. A complementary and potentially more versatile alternative would use ubiquitous alkenes, which are commonly encountered in petrochemical feedstocks and synthetic intermediates, as starting materials (*7*). Although aminofunctionalization reactions, such as hydroamination (*8*–*13*) and aminohydroxylation (*14*–*17*) of alkenes and alkene azidation (*18*–*21*), have played an important role in this area, these reactions are generally limited by the need to protect (and later deprotect) the nitrogen, activate the olefin with an aryl substituent, or generate hazardous azide intermediates (Fig. 1A). Falck and Kürti recently reported the preparation of a wide range of *NH*-aziridines via rhodium catalysis, a reaction that addresses some important challenges in this area (*22*, *23*). However, the application of this reaction is limited by the challenges associated with the subsequent regioselective opening of the aziridines to reveal the desirable primary amine building blocks, as well as the high cost of rhodium. Thus, the direct catalytic and regioselective synthesis of unprotected amine derivatives from unactivated alkenes remains an unsolved challenge in organic chemistry.

A catalytic method to introduce both the desired unprotected amine group, ideally at the more challenging primary position, and a versatile electrophile at the secondary position would prove particularly useful in the preparation of important compounds because of the complementary reactivity of these functional groups. Ideally, these products could be accessed by using a simple protocol, inexpensive reagents, and an Earth-abundant metal catalyst. Unprotected 2-chloroalkylamines would, in principle, fulfill the role of amphoteric aminated building blocks. These seemingly unstable molecules have rarely been reported in the literature (*24*), and the few viable synthetic methods to access them have generally relied on the direct chlorination of amino alcohols (*25*) or the ring opening of *NH*-aziridines (*26*). Alternatively, derivatives protected by a tosyl (Ts) group can be occasionally accessed

[1]Max-Planck-Institut für Kohlenforschung, Mülheim an der Ruhr, Germany. [2]ETH Zürich, Zürich, Switzerland.
*These authors contributed equally to this work.
†These authors contributed equally to this work.
‡Corresponding author. Email: morandib@ethz.ch

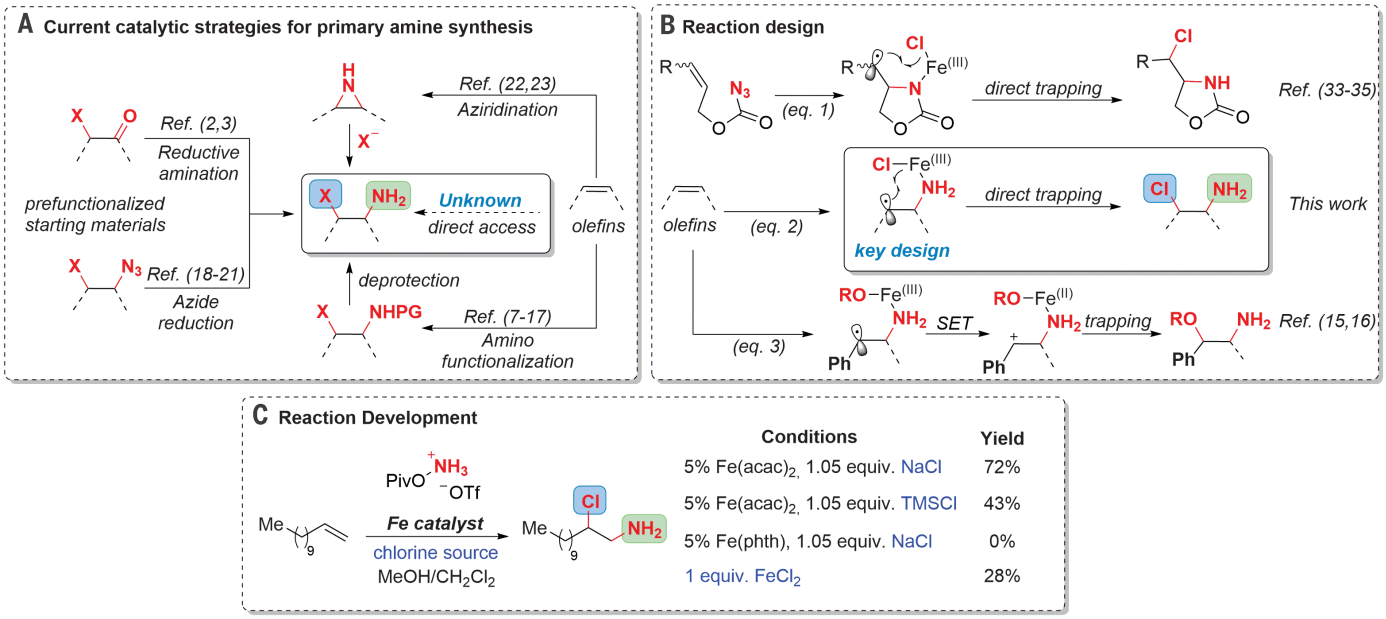


Fig. 1. Development of a strategy for the synthesis of a broad class of unprotected amines. (**A**) Traditional approach for primary amine synthesis. X, polar group or H; PG, protecting group. (**B**) Our design for direct radical trapping. R, any group; Ph, phenyl; SET, single electron transfer. (**C**) Reaction development (selected conditions). Piv, pivaloyl; Tf, trifluoromethanesulfonyl; TMSCl, trimethylsilyl chloride; phth, phthalimide.

from alkenes by using stoichiometric protocols, but these reactions are limited by the challenging selective cleavage of the Ts group in the presence of the chloride and the lack of regioselectivity, as well as poor scope and yield (*27*–*32*). An intramolecular iron-catalyzed aminochlorination has been reported by Bach *et al.*, though the products are chlorosubstituted oxazolidinones rather than unprotected 2-chloroalkylamines (*33*–*35*).

In iron-catalyzed amination reactions of alkenes more generally (*11*, *15*–*17*, *36*–*41*), the use of specific substituents (e.g., Ts) on the nitrogen atom to control reactivity and the high inefficiency when using unactivated olefins remain limiting factors. To collectively address these issues, we reasoned that the use of chloride as a nucleophile in electrophilic amination reactions would not only enable access to versatile 2-chloroalkylamines but also provide a strategy to address the lack of reactivity observed with traditionally unreactive alkenes. Related (*15*, *16*) iron-catalyzed aminohydroxylation reactions and etherifications have a limited substrate scope because they rely on a mechanism that involves a single electron transfer–mediated oxidation of a carbon-centered radical to a carbocation, a process that becomes very challenging when nonbenzylic radicals are generated as intermediates (Fig. 1B, Eq. 3). On the basis of Bach's precedent (Fig. 1B, Eq. 1) and the well-known propensity of the chlorine atom to undergo homolytic substitution reactions (*42*), we hypothesized that the putative unstable radical intermediate generated through the attack of a nitrogen-centered radical on a C=C double bond could instead rapidly undergo chlorine atom transfer from the amine-bound iron(III) complex to form the desired product and regenerate the iron(II) catalyst (Fig. 1B, Eq. 2). In this strategy, the facile homolytic substitution reaction of the carbon radical with the iron-bound chlorine atom, a pathway not easily accessible with oxygen-containing ligands, provides a direct route to product formation that does not involve any additional single electron transfer step or cationic intermediates. Consequently, a broader palette of alkenes should be amenable to this new process.

Herein, we report the iron-catalyzed aminochlorination of a wide range of conjugated and nonconjugated alkenes, including densely functionalized substrates. The reaction proceeds with excellent regioselectivity for the linear amine [the minor branched isomer could not be detected by proton nuclear magnetic resonance (H-NMR) spectroscopy] and leads to the direct formation of unprotected 2-chloroalkylamines under operationally simple conditions.

Initial proof-of-concept experiments were conducted under air with a stoichiometric amount of $FeCl_2$ added to a solution of 1-dodecene (a traditionally challenging substrate in electrophilic amination reactions) in the presence of a variety of hydroxylamine-derived reagents (Fig. 1C and table S1). A substantial conversion to the product was observed, albeit in low yields because of the product's instability in the presence of excess Lewis acidic iron salts. We then explored different chloride sources to move away from the use of stoichiometric amounts of iron salts. Although trimethylsilyl chloride, the source used by Bach *et al.* (*33*–*35*), gave moderate yields of product, we discovered that the much more benign and easy-to-handle sodium chloride gave superior results. We believe that the use of a mild, non–Lewis acidic source of chloride is crucial to prevent the undesired decomposition of the relatively unstable aminochlorinated product. Optimal conditions for the reaction used inexpensive iron(II) acetylacetonate [$Fe(acac)_2$] as a catalyst, simple sodium chloride as a chloride source, and a hydroxylamine-derived reagent as an amine source at room temperature in a methanol-dichloromethane mixture. The 2-chloroalkylamine was isolated in 72% yield with no detected alternative regioisomers. The aminating reagent

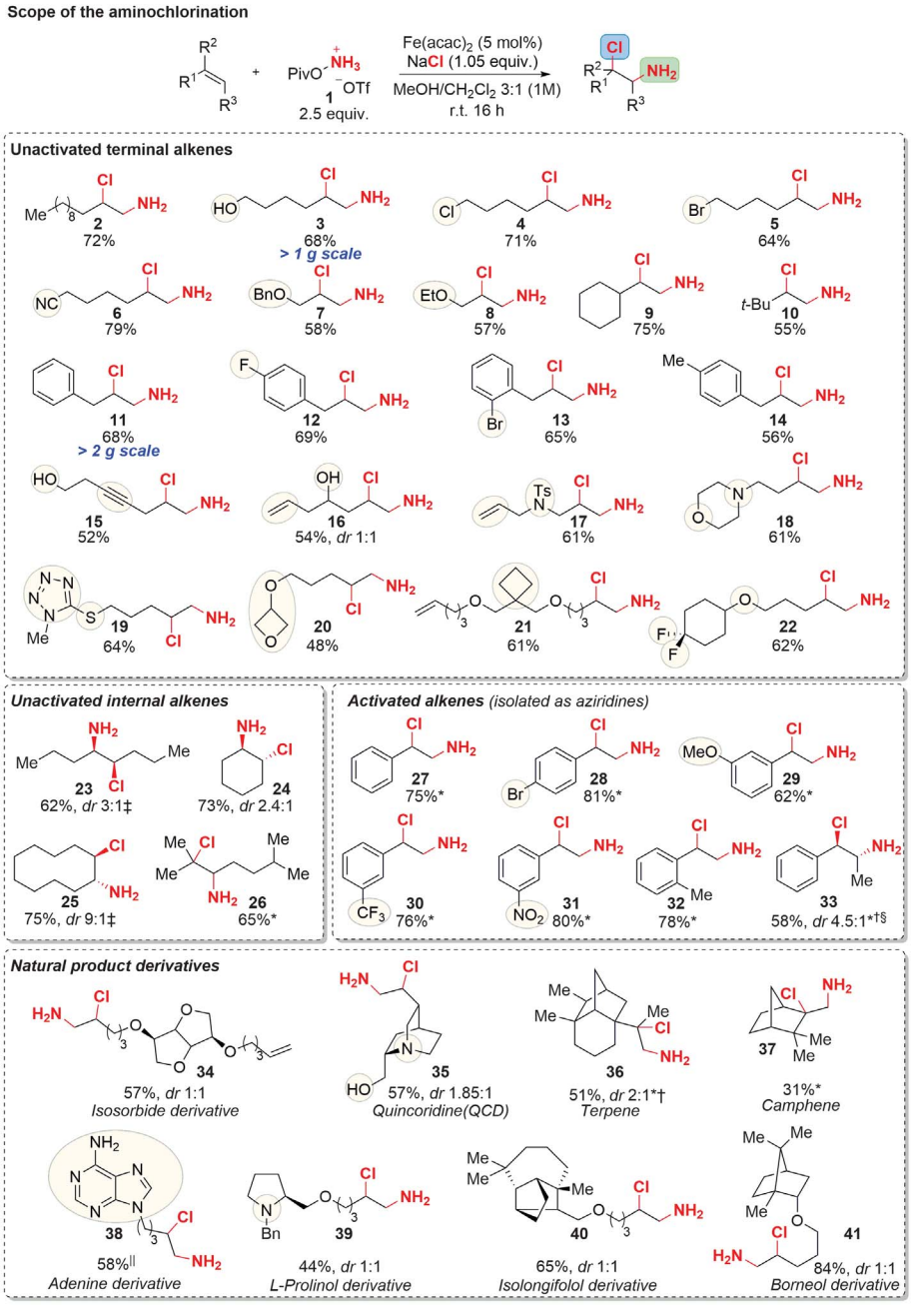


Fig. 2. Scope of the reaction with activated and unactivated alkenes. All yields, unless otherwise noted, are isolated yields of the 2-chloroalkylamine products. *Yields refer to the isolated aziridine products after basic workup. †Diastereomeric ratio of the isolated aziridine products. ‡(*E*)-Alkene isomer used as substrate. §(*Z*)-alkene isomer used as substrate. ||^{1}H-NMR yield; see the supplementary materials for isolation and characterization. *dr*, diastereomeric ratio; r.t., room temperature.

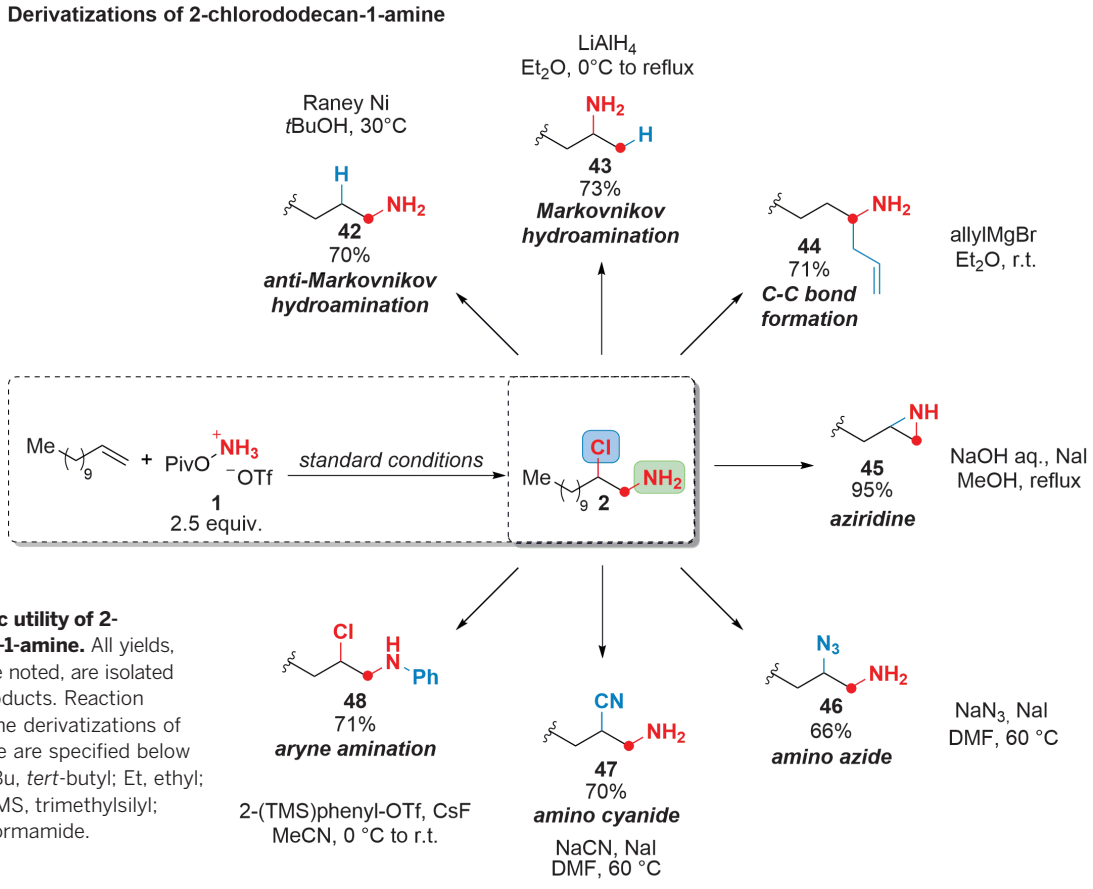


Fig. 3. Synthetic utility of 2-chlorododecan-1-amine. All yields, unless otherwise noted, are isolated yields of the products. Reaction conditions for the derivatizations of 2-chloro-1-amine are specified below the products. *t*Bu, *tert*-butyl; Et, ethyl; aq., aqueous; TMS, trimethylsilyl; DMF, dimethylformamide.

is thermally stable up to a temperature of 160°C, as demonstrated by differential scanning calorimetry (DSC) analysis (fig. S2).

We next explored the alkene substrate scope of this reaction, including two gram-scale demonstrations. A wide range of terminal alkenes, some containing unprotected polar groups that are traditionally problematic in transition metal catalysis (e.g., cyano substrate **6** and hydroxy substrates **3**, **15**, and **16**), gave the corresponding products in good yields (Fig. 2). Basic tertiary nitrogen centers (**35**, **18**, **38**, and **39**) were also tolerated, an important prerequisite for the application of any methodology for the discovery of bioactive compounds. Heterocycles commonly used in drug discovery, such as tetrazole (**19**), oxetane (**20**), and purine (**38**), could be readily used, as could an alkyne (**15**). Furthermore, aromatic halides (**12** and **13**), which can serve as handles for further functionalization through cross-coupling, gave the desired product in good yields. A striking feature of the reaction was the exclusive formation of the monofunctionalized product even in cases where two alkenes were present in the starting material (**16**, **17**, and **34**); the absence of difunctionalization product in the crude reaction mixture suggests that the low solubility of the protonated amine product prevents overfunctionalization. The reaction is not limited to monosubstituted alkenes. Both 1,1 (**36** and **37**)- and 1,2 (**23** to **25**)-disubstituted alkenes and even a trisubstituted alkene (**26**) also gave the products in high yields with partial retention of stereochemistry. Styrenes (**27** to **33**) could also be used in this transformation; however, in this case, the instability of the benzylic chlorides prevented the isolation of the products. Instead, the corresponding aziridine was isolated directly after basic workup. With regard to the substrates that failed to produce synthetically useful yields of products (fig. S3), solubility issues encountered with cholesterol derivatives and the lack of chemoselectivity with polyolefin substrates account for most of the limitations observed thus far.

We next probed the range of densely functionalized substrates amenable to the reaction. A derivative of isosorbide (**34**), a renewable feedstock, as well as several structurally complex terpene derivatives (**36**, **37**, **40**, and **41**) were aminochlorinated in good yields. More importantly, nitrogen-based bioactive scaffolds, such as quinocoridine (**35**), prolinol (**39**), and even unprotected adenine (**38**), delivered the corresponding aminated products. Such substrates would arguably pose a great challenge to the vast majority of other transition metal–catalyzed amination reactions.

Unprotected 2-chloroalkylamines are rare compounds whose reactivity has not yet been studied extensively. Having a convenient method to access them, we sought to explore their reactivity and synthetic utility (Fig. 3). In particular, it was our goal to address current challenges in amination chemistry through the use of these amphoteric building blocks. The catalytic synthesis of primary amines from terminal olefins has been recognized as one of the top 10 challenges in catalysis (*43*). 2-Chloroalkylamines can easily be reduced to the corresponding amines (**42**), affording a practical solution to this challenge. Changing the reducing agent and solvent even enables access to the Markovnikov hydroamination product (**43**), a process that probably proceeds through in situ generation of an aziridine intermediate followed by reductive ring opening. Common nucleophiles, such as azide (**46**) and cyanide (**47**), led to a practical synthesis of the corresponding products, which are versatile building blocks for diamine and amino acid synthesis, respectively. The corresponding aziridines (**45**) can also be easily accessed upon simple treatment with a base, providing a convenient iron-based alternative to the rhodium method reported by Kürti and Falck. Surprising reactivity was observed with allyl magnesium bromide, as C–C bond formation occurred at the α position relative to the amino group, leading to useful unprotected homoallylic amines (**44**). This reaction likely proceeds through initial HCl elimination to form an imine intermediate that can subsequently be attacked by the Grignard reagent.

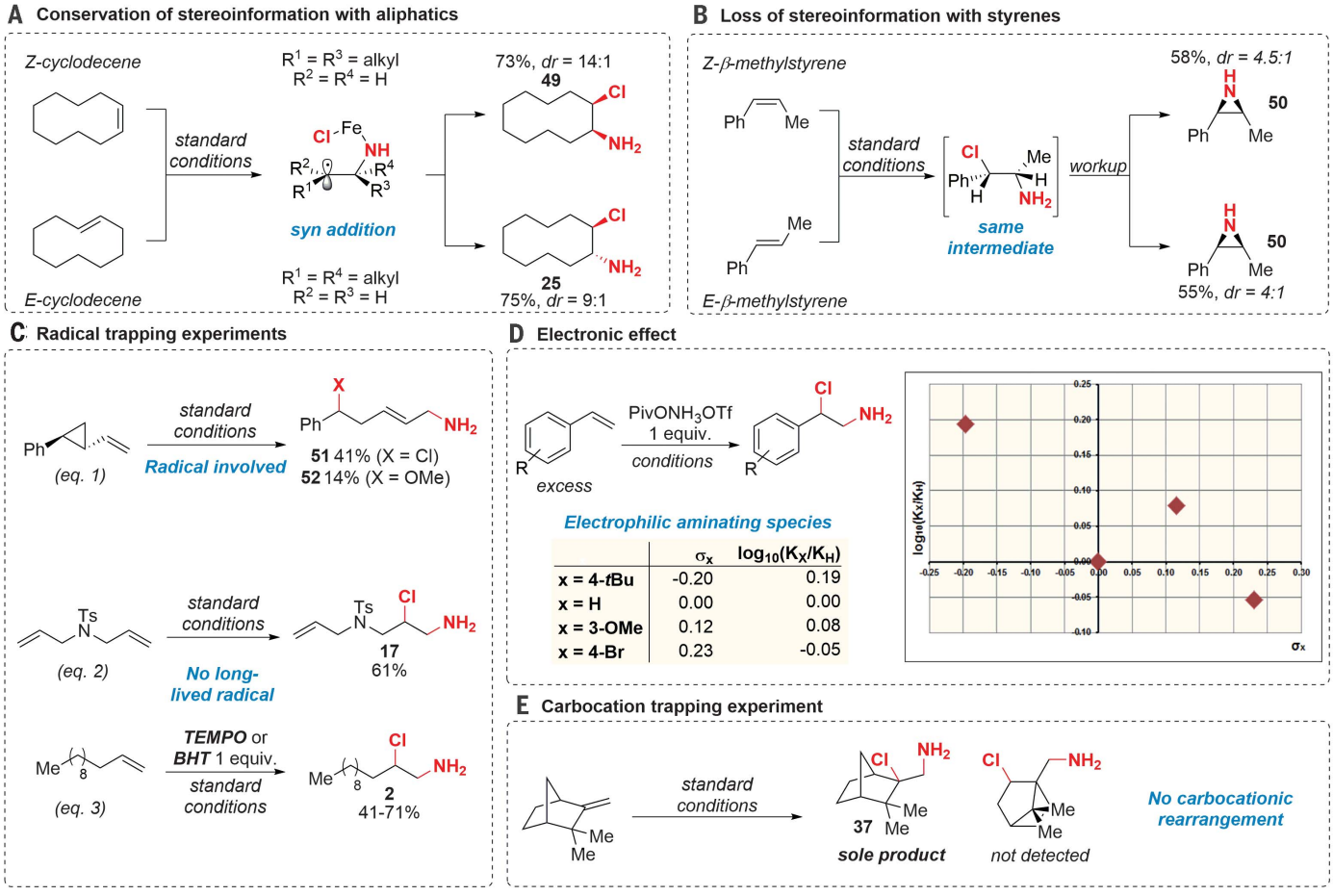


Fig. 4. Preliminary mechanistic experiments. (**A**) Conservation of stereoinformation with cyclodecene substrates indicating a syn addition mechanism. (**B**) Loss of stereoinformation with β-Me styrene compounds showing a stepwise mechanism with this class of substrates. (**C**) Radical experiments supporting the intermediacy of a short-lived radical intermediate. (**D**) Preliminary Hammett study supporting a mechanism involving an electrophilic amination mechanism. σ_x, Hammett constant; K_X/K_H, ratio of the two rates. (**E**) Lack of backbone rearrangement supporting the absence of carbocationic intermediates.

Preliminary studies were performed to support our initial mechanistic hypothesis (Fig. 1B). The use of *cis*- and *trans*-cyclodecene led to the isolation of the corresponding products (**49** and **25**) in good yields and with good overall conservation of stereochemistry (Fig. 4A). This result suggests an efficient formal syn delivery of the NH_2 and Cl groups. A similar experiment using both *cis*– and *trans*–β-methyl (β-Me) styrene resulted in a completely different outcome (Fig. 4B). The stereoinformation was entirely lost in this case, and both starting materials led to the same stereoisomer of the major product after conversion to the aziridines through stereoinvertive ring closing (*44*). This divergent outcome between the two classes of substrates is consistent with the proposed model (Fig. 1B), wherein an amine-coordinated iron complex rapidly delivers the chloride from the same face with nonstabilized aliphatic alkenes as substrates, whereas rapid oxidation of the radical intermediate to a stable benzylic cation occurs in styrenes, concomitant with stereochemical randomization.

We next aimed to support experimentally the postulated formation of the radical in the aliphatic alkene reaction. Experiments using one of the fastest radical clocks to detect the intermediacy of short-lived radical species supported the formation of radical intermediates, because ring opening of the cyclopropane could be clearly observed (Fig. 4C, Eq. 1). This result, combined with the high stereoretention of the reaction with aliphatic alkenes (Fig. 4A) and the lack of cyclization of another radical clock (Fig. 4C, Eq. 2), supports the formation of a short-lived carbon-centered radical intermediate. This conclusion was further supported by radical trapping experiments using TEMPO [(2,2,6,6-tetramethylpiperidin-1-yl)oxyl] and BHT (3,5-di-tert-4-butylhydroxytoluene), which did not substantially influence the reactivity (Fig. 4C, Eq. 3). On the basis of those experiments, we believe that the rate constant of the intramolecular chlorine atom transfer is very high, although it cannot exceed the known rate constant of the cyclopropyl radical clock (3×10^{11} s^{-1}) (*45*). A preliminary Hammett study showed a rate increase for electron-rich substrates, a result consistent with a mechanism involving the attack of highly electrophilic aminating species onto the alkene (Fig. 4D). Finally, we performed a simple experiment aiming to trap a possible carbocationic intermediate (Fig. 4E) through a rapid cationic rearrangement. In contrast to previous reports (*15*, *46*), no rearrangement was observed under our reaction conditions, supporting the operation of a distinct mechanistic pathway.

Collectively, these results are best explained by our initially proposed model and support the direct intramolecular trapping of a short-lived carbon-centered radical through intramolecular *syn*-chlorine atom transfer from an amine-coordinated iron(III)-Cl complex.

REFERENCES AND NOTES

1. L. Legnani, B. N. Bhawal, B. Morandi, *Synthesis* **49**, 776–789 (2017).
2. T. Gross, A. M. Seayad, M. Ahmad, M. Beller, *Org. Lett.* **4**, 2055–2058 (2002).
3. R. V. Jagadeesh *et al.*, *Science* **358**, 326–332 (2017).
4. C. Gunanathan, D. Milstein, *Angew. Chem. Int. Ed. Engl.* **47**, 8661–8664 (2008).

5. C. Defieber, M. A. Ariger, P. Moriel, E. M. Carreira, *Angew. Chem. Int. Ed. Engl.* **46**, 3139–3143 (2007).
6. M. J. Pouy, L. M. Stanley, J. F. Hartwig, *J. Am. Chem. Soc.* **131**, 11312–11313 (2009).
7. P. Dauban, B. Darses, A. Jarvis, in *Comprehensive Organic Synthesis*, P. Knochel, G. A. Molander, Eds. (Elsevier, ed. 2, 2014), chap. 7.19, pp. 538–604.
8. M. Beller, J. Seayad, A. Tillack, H. Jiao, *Angew. Chem. Int. Ed. Engl.* **43**, 3368–3398 (2004).
9. Y. Miki, K. Hirano, T. Satoh, M. Miura, *Angew. Chem. Int. Ed. Engl.* **52**, 10830–10834 (2013).
10. S. Zhu, N. Niljianskul, S. L. Buchwald, *J. Am. Chem. Soc.* **135**, 15746–15749 (2013).
11. J. Gui *et al.*, *Science* **348**, 886–891 (2015).
12. D. G. Kohler, S. N. Gockel, J. L. Kennemur, P. J. Waller, K. L. Hull, *Nat. Chem.* **10**, 333–340 (2018).
13. A. J. Musacchio *et al.*, *Science* **355**, 727–730 (2017).
14. M. A. Andersson, R. Epple, V. V. Fokin, K. B. Sharpless, *Angew. Chem. Int. Ed. Engl.* **41**, 472–475 (2002).
15. L. Legnani, B. Morandi, *Angew. Chem. Int. Ed. Engl.* **55**, 2248–2251 (2016).
16. D.-F. Lu, C.-L. Zhu, Z.-X. Jia, H. Xu, *J. Am. Chem. Soc.* **136**, 13186–13189 (2014).
17. K. S. Williamson, T. P. Yoon, *J. Am. Chem. Soc.* **134**, 12370–12373 (2012).
18. Y.-A. Yuan, D.-F. Lu, Y.-R. Chen, H. Xu, *Angew. Chem. Int. Ed. Engl.* **55**, 534–538 (2016).
19. N. Fu, G. S. Sauer, A. Saha, A. Loo, S. Lin, *Science* **357**, 575–579 (2017).
20. K. Shen, Q. Wang, *J. Am. Chem. Soc.* **139**, 13110–13116 (2017).
21. X. Geng, F. Lin, X. Wang, N. Jiao, *Org. Lett.* **19**, 4738–4741 (2017).
22. J. L. Jat *et al.*, *Science* **343**, 61–65 (2014).
23. Z. Ma, Z. Zhou, L. Kürti, *Angew. Chem. Int. Ed. Engl.* **56**, 9886–9890 (2017).
24. P. Kovacic, M. K. Lowery, K. W. Field, *Chem. Rev.* **70**, 639–665 (1970).
25. G. Berger *et al.*, *Tetrahedron Lett.* **54**, 545–548 (2013).
26. E. Leemans, S. Mangelinckx, N. De Kimpe, *Synlett* **2009**, 1265–1268 (2009).
27. S. R. Chemler, M. T. Bovino, *ACS Catal.* **3**, 1076–1091 (2013).
28. Y. Cai *et al.*, *J. Am. Chem. Soc.* **133**, 5636–5639 (2011).
29. M. T. Bovino, S. R. Chemler, *Angew. Chem. Int. Ed. Engl.* **51**, 3923–3927 (2012).
30. Y. Cai *et al.*, *Chem. Commun.)* **2013**, 8054–8056 (2013).
31. F. Minisci, R. Galli, M. Cecere, *Tetrahedron Lett.* **7**, 3163–3166 (1966).
32. F. Minisci, R. Galli, *Tetrahedron Lett.* **6**, 1679–1684 (1965).
33. T. Bach, B. Schlummer, K. Harms, *Chem. Commun.* **2000**, 287–288 (2000).
34. T. Bach, B. Schlummer, K. Harms, *Chemistry* **7**, 2581–2594 (2001).
35. H. Danielec, J. Klügge, B. Schlummer, T. Bach, *Synthesis* **3**, 551–556 (2006).
36. A. Correa, O. García Mancheño, C. Bolm, *Chem. Soc. Rev.* **37**, 1108–1117 (2008).
37. D.-F. Lu, C.-L. Zhu, J. D. Sears, H. Xu, *J. Am. Chem. Soc.* **138**, 11360–11367 (2016).
38. C.-L. Zhu, J.-S. Tian, Z.-Y. Gu, G.-W. Xing, H. Xu, *Chem. Sci.* **6**, 3044–3050 (2015).
39. J.-S. Tian, C.-L. Zhu, Y.-R. Chen, H. Xu, *Synthesis* **47**, 1709–1715 (2015).
40. E. R. King, E. T. Hennessy, T. A. Betley, *J. Am. Chem. Soc.* **133**, 4917–4923 (2011).
41. E. T. Hennessy, R. Y. Liu, D. A. Iovan, R. A. Duncan, T. A. Betley, *Chem. Sci.* **5**, 1526–1532 (2014).
42. C. H. Schiesser, L. M. Wild, *Tetrahedron* **52**, 13265–13314 (1996).
43. J. Haggin, *Chem. Eng. News* **71**, 23–27 (1993).
44. A. Cruz, I. I. Padilla-Martínez, E. V. García-Báez, *Tetrahedron Asymmetry* **21**, 909–913 (2010).
45. D. C. Nonhebel, *Chem. Soc. Rev.* **22**, 347–359 (1993).
46. R. R. Naredla, D. A. Klumpp, *Chem. Rev.* **113**, 6905–6948 (2013).

ACKNOWLEDGMENTS

We thank B. Bhawal for critical proofreading of this manuscript and our apprentice, S. Begoihn, for the synthesis of some compounds. We also thank A. Pommerin for the DSC measurements. We kindly thank B. List for sharing analytical equipment, and the MS department of the MPI for technical assistance. **Funding:** Generous funding from the Max Planck Society, the Max-Planck-Institut für Kohlenforschung, the European Research Commission (grant agreement ShuttleCat 757608), the ETH Zürich, and the Fonds der Chemischen Industrie is acknowledged. **Author contributions:** B.M., L.L., and G.P.-C. designed the project. L.L., G.P.-C., T.D., and S.W. performed the experiments and collected the data. All authors analyzed the data and contributed to the writing of the manuscript. L.L. and G.P.-C. contributed equally as first authors. T.D. and S.W. contributed equally as third authors. **Competing interests:** We declare no competing financial interests. **Data and materials availability:** All experimental data are available in the supplementary materials. X-ray structural data can be accessed free of charge from the Cambridge Crystallographic Data Centre under reference numbers CCDC-1863481 to CCDC-1863483.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6413/434/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S6
Tables S1 to S4
References (*47*–*56*)

21 February 2018; resubmitted 6 July 2018
Accepted 4 September 2018
10.1126/science.aat3863

OPTICS

Synchrotron radiation from an accelerating light pulse

M. Henstridge[1,2], C. Pfeiffer[1,3], D. Wang[1,4], A. Boltasseva[1,4], V. M. Shalaev[1,4], A. Grbic[1,3], R. Merlin[1,2]*

Synchrotron radiation—namely, electromagnetic radiation produced by charges moving in a curved path—is regularly generated at large-scale facilities where giga–electron volt electrons move along kilometer-long circular paths. We use a metasurface to bend light and demonstrate synchrotron radiation produced by a subpicosecond pulse, which moves along a circular arc of radius 100 micrometers inside a nonlinear crystal. The emitted radiation, in the terahertz frequency range, results from the nonlinear polarization induced by the pulse. The generation of synchrotron radiation from a pulse revolving about a circular trajectory holds promise for the development of on-chip terahertz sources.

Synchrotron radiation (that is, the emission from a relativistic charge moving along a circular trajectory) was first observed in 1947 (*1*). Until the 1960s, synchrotrons were used to accelerate charged particles exclusively for experiments in particle physics, and radiation losses were studied primarily because they were an impediment to achieving high energies (*2*–*4*). Presently, large-scale synchrotrons that produce high-intensity x-rays are used in a wide range of scientific studies. Synchrotron radiation has also been produced at radio frequencies from accelerating polarization currents created by capacitor arrays (*5*) and is commonly observed in astrophysical phenomena such as radio galaxies and supernovae (*6*).

Here we present the observation of synchrotron radiation resulting from the acceleration of a light pulse (*7*–*11*). Our work distinguishes itself from previous studies involving accelerated beams in nonlinear media (*12*–*14*) in that the emphasis is placed on the identification of features exclusively associated with the centripetal acceleration of the primary beam. A metasurface (*15*–*20*) was used to guide a pulse of 800-nm central wavelength and 100-fs duration along a 100-μm–scale circular arc inside a lithium tantalate ($LiTaO_3$) crystal. As the pulse mixes with itself through the second-order susceptibility of the crystal, the accelerating light pulse generates both second-harmonic and difference-frequency nonlinear polarizations. As for a charge moving under the influence of a centripetal force in vacuum, interference effects lead to a nonlinear polarization, which moves along a circular path and emits synchrotron radiation, as if acted upon a fictitious force (*21*). We measure the emission from the difference-frequency polarization, which is in the terahertz range, over a scale of 100 μm.

In $LiTaO_3$, the nonlinear polarization generated by a 800-nm pulse propagates at approximately three times the phase velocity of light at terahertz frequencies, below those of the lowest-lying transverse optical (TO) phonons of A_1 and E symmetry (*22*). As a result, synchrotron radiation at these frequencies interweaves strongly with Cherenkov radiation (*23*–*26*), and as such, its spectrum has features that are fundamentally different from those of charges circulating at subluminal speeds. Because the polarization field defines the local dipole density, the general properties of the radiation produced by our pulses relate to those of emission resulting from the motion of a single dipole or, more simply, from a charge q that moves at a constant speed v about a circular path. Assuming that the charge makes an infinite number of revolutions, the power emitted into the mth harmonic of the revolution frequency, $\omega_0 = v/R$, is given by (*26*)

$$P_m = \frac{q^2 v m \omega_0^2}{c^2}\left[2J'_{2m}(u) + (1 - c^2/n^2v^2)\int_0^u J_{2m}(x)dx\right] \quad (1)$$

where R is the radius of the circle; J_{2m} and J'_{2m} are, respectively, the Bessel function of order $2m$ and its derivative; c is the speed of light in vacuum; n is the index of refraction at the mth harmonic and $u = 2mvn/c$. This expression is valid both for $v < c/n$ and $v > c/n$. For light polarized parallel to the optical axis of $LiTaO_3$, $n = 6.3$ at frequencies considerably below that of the lowest-lying A_1-type TO phonon, at 6.0 THz (*22*), whereas the group index at 800 nm is $n_g = 2.2$ (*27*). Figure 1 shows the calculated discrete power spectra for emission from a charge at both subluminal and superluminal speeds in a medium with constant $n = 6.3$. The radiated power increases with frequency when $v > c/n$, whereas for $v < c/n$, the emission spectrum exhibits a cutoff at a critical frequency $\omega_C = 3\omega_0/2(1 - v^2n^2/c^2)^{3/2}$ beyond which the radiation emitted is negligible (this is more clearly seen on a linear scale) (*28*). In our experiments, the highest synchrotron frequency attainable is determined by the bandwidth of the pump

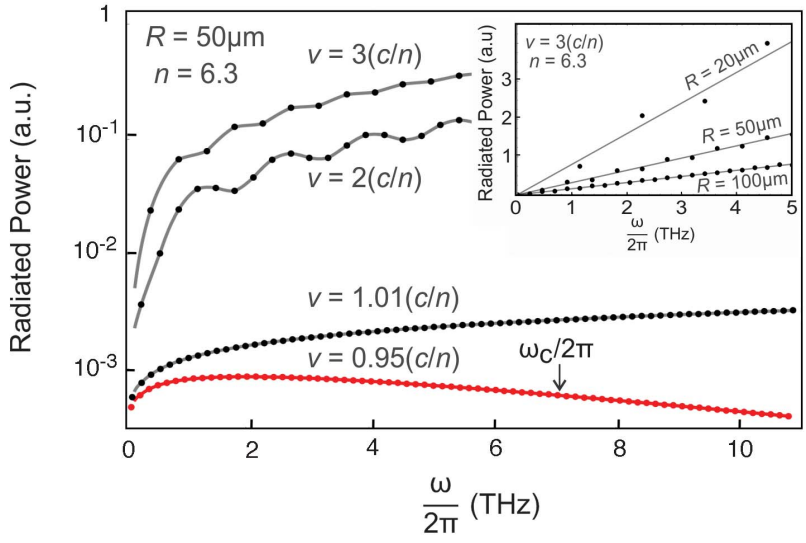


Fig. 1. Radiated power: Comparison between subluminal and superluminal speeds. Calculated power radiated by a charge moving in a circle of radius $R = 50$ μm, at speeds below and above c/n, in a medium with constant $n = 6.3$ (note the logarithmic scale) is shown. Emission is at integer multiples of the fundamental frequency $\omega_0 = v/r$. At subluminal speeds, the emitted power drops rapidly for frequencies greater than the critical angular frequency ω_C. (**Inset**) Results for $v = 3c/n$ and three different radii (linear scale). For speeds $v >> c/n$, the power emitted into harmonics behaves linearly with slope given by $(q^2v/c^2)(1 - c^2/v^2n^2)\omega_0$. The curves and lines are guides for the eye. a.u., arbitrary units.

[1]Center for Photonics and Multiscale Nanomaterials, University of Michigan, Ann Arbor, 48109, USA. [2]Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA. [3]Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109, USA. [4]School of Electrical and Computer Engineering and Birck Nanotechnology Center, Purdue University, West Lafayette, IN 47907, USA.
*Corresponding author. Email: merlin@umich.edu

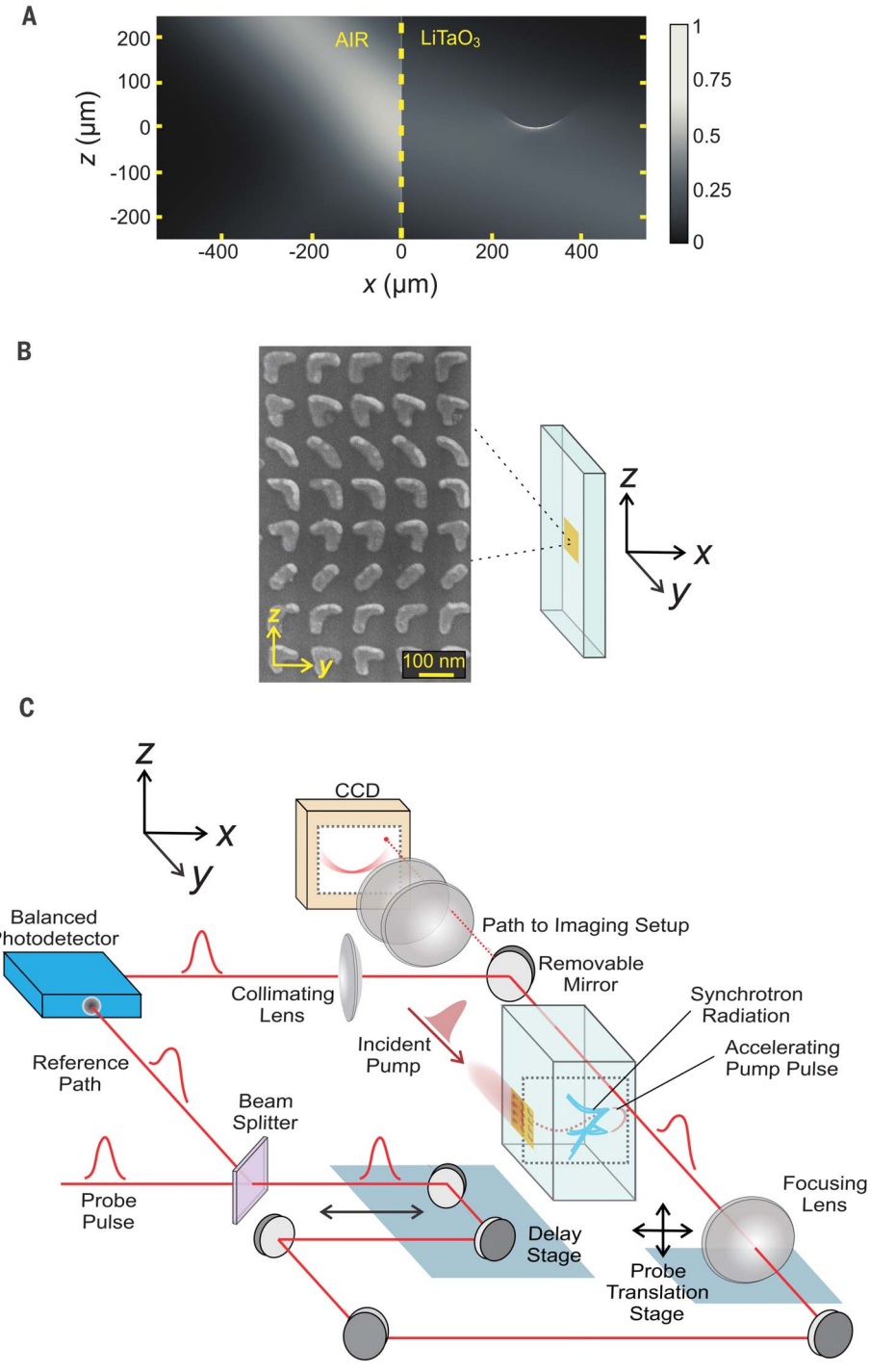


Fig. 2. Accelerated beam, metasurface, and experimental setup. (**A**) Two-dimensional plot showing the intensity of the incident and elastically scattered field from the metasurface, located at $x = 0$. The reflected beam intensity is roughly five times smaller than that of the incident beam. (**B**) Scanning electron microscope image of a section of the nanoantenna array fabricated onto the $LiTaO_3$ substrate. To facilitate the imaging of the accelerating beam, the leading edge of the metasurface was positioned ≈1 mm away from the edge of the substrate. The array was aligned to obtain an accelerating beam in the *xz* plane, in accordance with the 2D plot shown in (A). The pattern repeats periodically along the *y* axis; the width of the accelerating beam parallel to this axis is determined by the width of the incident beam. (**C**) Schematic of the orthogonal pump-probe setup. The output of a measurement set is an image sequence showing the time evolution of the time derivative of the pump-pulse intensity and the terahertz synchrotron radiation field. The dotted curve inside the crystal denotes the trajectory of the accelerating pump pulse. CCD, charge-coupled device.

pulses. For speeds considerably greater than c/n, the radiated power increases substantially. For example, the power radiated at a given frequency for $v = 3c/n$ is several orders of magnitude larger than that at a nearby frequency for $v = 0.95c/n$. We emphasize the facts that these results assume a frequency-independent refractive index and that the spectra consist of discrete lines only if the motion is strictly periodic; a source undergoing a few cycles or a fraction of a cycle emits instead a continuous spectrum. This applies to conventional synchrotron radiation for which the source is a single charge, as well as to a nonlinear polarization moving along a circular path.

Calculated spectra for an electron revolving in a medium with constant $n = 6.3$ and $v = 3c/n$ for three different radii are shown in the inset of Fig. 1. The values for the electron velocity and light speed are similar to those in our experiments on $LiTaO_3$. The fact that the radiation involves a few discrete frequencies suggests that light-produced synchrotron terahertz radiation could be useful as a means to obtain on-chip continuous-wave terahertz sources. Furthermore, synchrotron emission from a nonlinear accelerating polarization can serve as a new method for generating coherent light at any frequency and in any medium at which conventional phase-matching with the nonlinear polarization is poor, such as at frequencies immediately above the Reststrahlen band in crystals such as ZnTe or GaP, or at second-harmonic frequencies. Within this context, we note a report of unusually phase-matched second-harmonic generation using Bessel beams (*29*).

We designed a plasmonic nanoantenna metasurface (*17*, *18*) to accelerate light of wavelength 800 nm along a circular arc of radius of curvature equal to 100 μm, inside the $LiTaO_3$ crystal. Unlike conventional methods for accelerating light, which rely on spatial light modulators (*30*), metasurfaces have the ability to bend light at arbitrarily large angles from the normal to a planar interface and thereby produce trajectories of very small radii inside a material (*31*). The nanoantennas scatter roughly 5% of the incident power into the accelerating beam, which is cross-polarized relative to the incident polarization. Approximately 50% of the incident beam power is transmitted unperturbed and ~20% is reflected. To reduce the spatial overlap between the unperturbed and accelerating beams, the nanoantennas were arranged to produce an accelerating beam when illuminated at a 45° angle of incidence. The design process is described in detail in (*31*) and summarized in the supplementary materials (*32*). A two-dimensional (2D) plot showing the calculated incident and scattered intensities for continuous-wave illumination at 800 nm is shown in Fig. 2A. Calculations of the accelerating pulse produced by the metasurface when illuminated with a pulse from an ultrafast laser are presented in (*32*).

The 500-μm–by–500-μm metasurface was fabricated onto a 1-mm-thick $LiTaO_3$ substrate by using electron-beam lithography and lift-off techniques. Prior to fabrication, all four of the 1-mm-wide sides of the substrate were polished to optical quality to eliminate Rayleigh scattering. Details are discussed in (*32*). A schematic describing the positioning of the metasurface and a scanning electron microscope image of the nanoantennas are shown in Fig. 2B.

The accelerating pulse was generated by illuminating the metasurface with a pulse of 800-nm central wavelength and duration of 100 fs from a regeneratively amplified Ti:sapphire oscillator operating at the repetition rate of 250 kHz. We performed a spatially resolved pump-probe experiment in the differential transmission geometry to detect the terahertz radiation generated by the accelerating pulse. The experimental setup is shown in Fig. 2C. The probe pulse is polarized parallel to the optical (z) axis while the electric field of the accelerating pump lies in the xz plane and rotates slightly as the pulse propagates. The component of the nonlinear polarization, $\mathbf{P}_{NL}$, relevant to our experiments is that along the optical axis $\mathbf{P}_{NL}.\hat{z} = \chi^{(2)}_{333}|\mathbf{E}_P.\hat{z}|^2$; here, $\mathbf{E}_P$ is the field of the pump, and $\chi^{(2)}_{333}$ is the corresponding element of the second-order nonlinear susceptibility. The probe beam directed to the sample was focused to a spot with a full width at half maximum (FWHM) of 10 μm in the plane of the accelerating beam trajectory. The probe's focusing lens was mounted onto a motorized translation stage, allowing for the translation of the focal spot throughout the acceleration plane. To calibrate the position of the probe, the accelerating and focused probe beams were first imaged together using a standard imaging setup. The transmitted probe was then directed to a balanced photodetector at which its average intensity was subtracted from that of the reference beam. The focal spot was translated throughout a 400-μm–by–400-μm area of the accelerating beam plane in steps of 10 μm. At each position, the arrival time of the probe pulse was delayed in steps of 53.4 fs, and the corresponding differential intensity trace was recorded. The output of a completed measurement is an image sequence showing the propagation of the time-derivatives of the 800-nm pump-pulse intensity and terahertz synchrotron field. We did not measure the power spectrum of the terahertz radiation. As noted earlier and as opposed to the discrete emission resulting from repeated rotations (Fig. 1), we expect it to be a continuum because the accelerating pulse moves just a fraction of one cycle. Further details concerning the experiment and the processing of the data are presented in the supplementary materials (*32*).

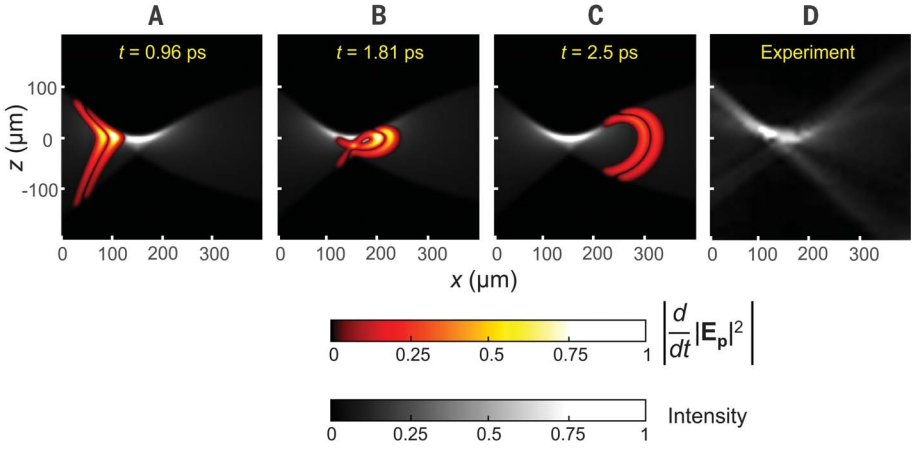


Fig. 3. Accelerating pump pulse. (**A** to **C**) Two-dimensional plots showing the calculated magnitude of the time derivative of the pump-pulse intensity, $|\mathbf{E}_P|^2$, and the intensity for continuous-wave excitation at 800 nm. To compare with the experimental data, results were numerically convolved with a Gaussian function with 10-μm FWHM. (**D**) Image of the accelerating beam trajectory obtained by adding the series of experimental traces for $|d|\mathbf{E}_P|^2/dt|$.

Figure 3 displays calculated 2D plots of the accelerating pump pulse alone, at different positions along the beam trajectory, as well as an experimental image of the accelerating beam trajectory obtained from the pump-probe measurements. The main results of this work are presented in Fig. 4. Panels A to D show plots of the measured time derivatives of the terahertz electric field and pump intensity for four different arrival times of the probe. In all cases, the unperturbed incident pulse, which passes through the metasurface, was digitally removed. Consistent with results reported in (*33*), this pulse generates a negligible amount of terahertz radiation at its edges, which was also removed from the data. Panels E to H of Fig. 4 show the corresponding theoretical simulations. For comparison purposes, results of calculations for the radiation emitted by a single dipole are displayed in panels I to L.

The theoretical difference-frequency synchrotron electric field was obtained by numerically solving the following scalar, inhomogeneous wave equation for each terahertz frequency component

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial z^2}+\frac{\Omega^2}{c^2}\varepsilon(\Omega)\right)E_S = \frac{1}{c^2}\frac{\partial^2 P_{NL}}{\partial t^2}\bigg|_\Omega \approx \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\left(\chi^{(2)}_{333}|\mathbf{E}_P|^2\right)\bigg|_\Omega \quad (2)$$

Here, Ω is the terahertz frequency and $f(t)|_\Omega$ denotes the Fourier transform of $f(t)$ at Ω. E_S (x, z, Ω) and P_{NL} are the z component of, respectively, the synchrotron electric field and the nonlinear polarization; $\mathbf{E}_P$ is the electric field of the 800-nm accelerating pump pulse; and $\varepsilon(\Omega)$ is the far-infrared permittivity of $LiTaO_3$ for radiation polarized parallel to the optical axis (*32*). For simplicity, we assumed that the polarization of the electric field is normal to the accelerating beam trajectory; that is, we ignored effects due to the rotation of the accelerating pump polarization. The theoretical plots in Fig. 4, E to H, were generated by taking the time derivatives of the solution to Eq. 2 and the calculated accelerating pulse intensity. The 2D matrix for each plot was then convolved with a Gaussian function of 10-μm FWHM to account for the size of the probe's focal spot. The complete image sequence showing the time evolution of the pump pulse and synchrotron field is presented in (*32*).

The simulations reproduce most of the features observed in the experiments, and there is reasonably good quantitative agreement between the two sets of results shown, respectively, in panels E to H and A to D of Fig. 4. Consistent with the optical properties of $LiTaO_3$, the 800-nm pulse propagates at a speed approximately three times that for the terahertz synchrotron field. As mentioned earlier, the superluminal nature of the propagation leads to new features in the radiation pattern, which distinguish themselves from those of ordinary, subluminal synchrotron

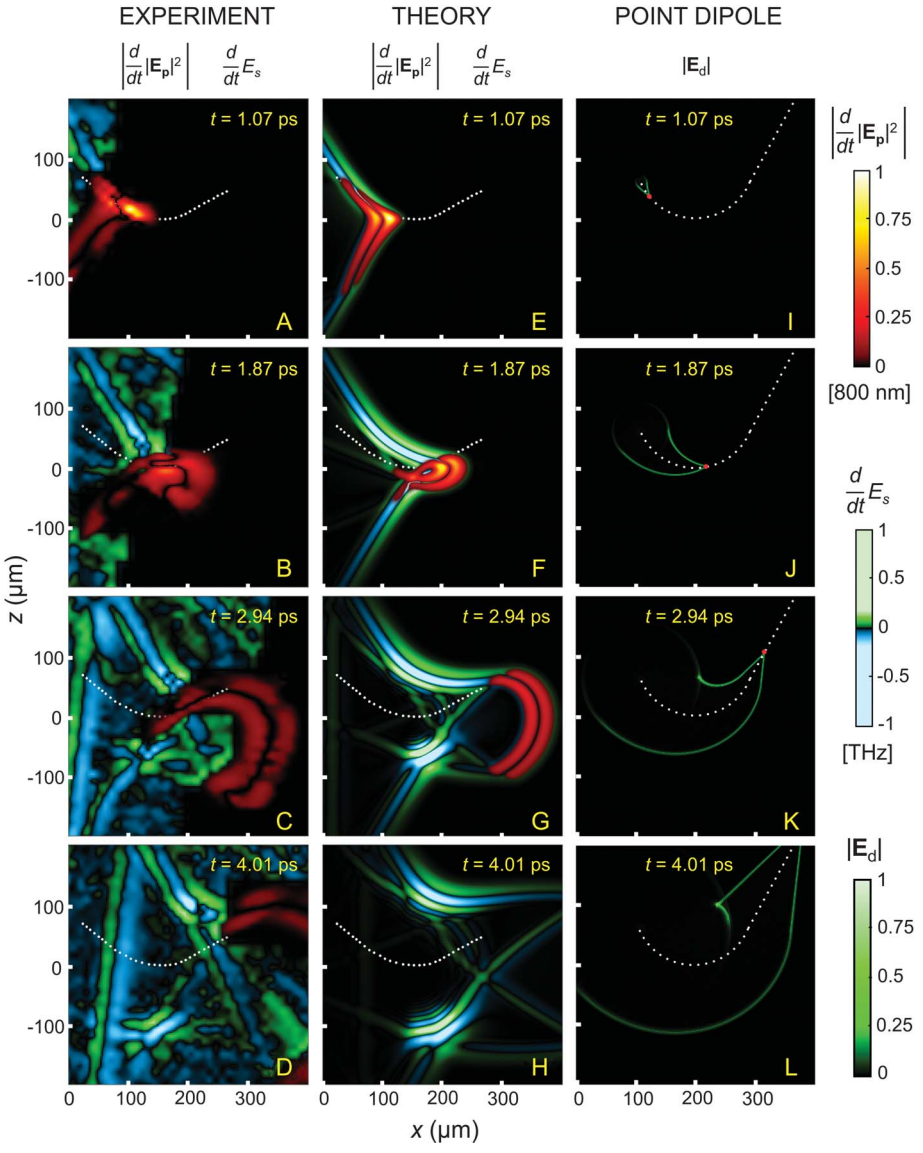


Fig. 4. Synchrotron radiation from a superluminal accelerating pulse. (**A** to **D**) Two-dimensional plots showing results of the orthogonal pump-probe measurements. The signal proportional to the time-derivatives of the pump intensity $|\mathbf{E}_P|^2$ and that of the terahertz field, E_S, were isolated and plotted on separate scales. (**E** to **H**) Theoretical $|d|\mathbf{E}_P|^2/dt|$ and dE_S/dt. The dashed white curve is the trajectory derived from calculations of the pump intensity. (**I** to **L**) Radiation emitted by a point dipole with speed 2.86 times faster than that of light (this value is the ratio between the zero-frequency refractive index of $LiTaO_3$ and the group index at 800 nm). The dipole, indicated by the red dot, propagates first along a circular arc subtended by an angle of $3\pi/4$ and then follows a linear path for $t > 2.6$ ps. The origin of time and the parameters of the dipole trajectory were chosen to provide a good qualitative comparison with the pulse propagation data; the dipole emission associated with the linear part of the motion simulates that of the pump pulse upon its diffraction at $t \gtrsim 2.94$ ps. The pulse emission patterns at $t = 1.07$ closely resemble that of conventional Cherenkov radiation. The observation of two Cherenkov-related curved branches, prominently observed for $t \geq 1.87$ ps in all panels, as well as the cusp in the inner radiation branch at $t = 2.94$ and 4.01 ps, are features characteristic of superluminal synchrotron radiation.

radiation. Specifically, both the theoretical and measured terahertz fields show the presence of two (inner and outer) radiation branches as opposed to a single one (*28*). As the early time panels A and E show, these branches meet in a form that strongly resembles the characteristic shock wave of Cherenkov radiation. By the time the pulse passes through the midpoint of the circular trajectory, the fields have developed prominent curvatures in both branches, which are readily apparent for $t \geq 1.87$ ps. Such a distorted, curvilinear Cherenkov pattern is also evident in the single-dipole results. Another key feature of superluminal speeds is the formation of a cusp in the inner radiation branch (*34*, *35*), which is clearly observed in the calculations for the single dipole. Note that the cusp is well developed by the time the dipole makes one-half of a revolution. The 800-nm pulse in the experiment completes between one-quarter and one-half of a revolution. Consistent with the single-dipole data, the motion of our pulses produces a small cusp in the inner radiation branch that can be seen in Fig. 4 at $t = 2.94$ ps near $(x, z) = (175\ \mu m, 75\ \mu m)$ and at $t = 4.01$ ps near $(x, z) = (200\ \mu m, 100\ \mu m)$.

REFERENCES AND NOTES

1. F. R. Elder, A. M. Gurewitsch, R. V. Langmuir, H. C. Pollock, *Phys. Rev.* **71**, 829–830 (1947).
2. D. Iwanenko, I. Pomeranchuk, *Phys. Rev.* **65**, 343 (1944).
3. J. P. Blewett, *Phys. Rev.* **69**, 87–95 (1946).
4. J. P. Blewett, *J. Synchrotron Radiat.* **5**, 135–139 (1998).
5. A. Ardavan *et al.*, *J. Appl. Phys.* **96**, 4614 (2004); **96**, 7760(E) (2004).
6. V. L. Ginzburg, S. I. Syrovatskii, *Annu. Rev. Astron. Astrophys.* **3**, 297–350 (1965).
7. G. A. Siviloglou, D. N. Christodoulides, *Opt. Lett.* **32**, 979–981 (2007).
8. G. A. Siviloglou, J. Broky, A. Dogariu, D. N. Christodoulides, *Phys. Rev. Lett.* **99**, 213901 (2007).
9. I. Kaminer, R. Bekenstein, J. Nemirovsky, M. Segev, *Phys. Rev. Lett.* **108**, 163901 (2012).
10. P. Zhang *et al.*, *Opt. Lett.* **37**, 2820–2822 (2012).
11. F. Courvoisier *et al.*, *Opt. Lett.* **37**, 1736–1738 (2012).
12. Y. Hu *et al.*, *Sci. Rep.* **7**, 8695 (2017).
13. I. Dolev, I. Kaminer, A. Shapira, M. Segev, A. Arie, *Phys. Rev. Lett.* **108**, 113903 (2012).
14. P. Polynkin, M. Kolesik, J. V. Moloney, G. A. Siviloglou, D. N. Christodoulides, *Science* **324**, 229–232 (2009).
15. R. Merlin, *Science* **317**, 927–929 (2007).
16. A. Grbic, L. Jiang, R. Merlin, *Science* **320**, 511–513 (2008).
17. N. Yu *et al.*, *Science* **334**, 333–337 (2011).
18. X. Ni, N. K. Emani, A. V. Kildishev, A. Boltasseva, V. M. Shalaev, *Science* **335**, 427 (2012).
19. D. Lin, P. Fan, E. Hasman, M. L. Brongersma, *Science* **345**, 298–302 (2014).
20. R. C. Devlin, M. Khorasaninejad, W. T. Chen, J. Oh, F. Capasso, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10473–10478 (2016).
21. I. Kaminer, J. Nemirovsky, M. Rechtsman, R. Bekenstein, M. Segev, *Nat. Phys.* **11**, 261–267 (2015).
22. A. S. Barker Jr., A. A. Ballman, J. A. Ditzenberger, *Phys. Rev. B* **2**, 4233–4239 (1970).
23. V. L. Ginzburg, *Phys. Usp.* (Engl. transl.) **39**, 973 (1996). [translation from *Usp. Fiziol. Nauk* **166**, 1033 (1996)].
24. V. L. Ginzburg, V. Ya. Éĭdman, *Sov. Phys. JETP* **36**, 1300 (1959).
25. G. N. Afanasiev, *Vavilov-Cherenkov and Synchrotron Radiation: Foundations and Applications* (Kluwer, 2005).
26. V. P. Zrelov, in *Cherenkov Radiation in High-Energy Physics* (Israel Program for Scientific Translations, 1970), p. 31.
27. J. K. Wahlstrand, R. Merlin, *Phys. Rev. B* **68**, 054301 (2003).
28. J. D. Jackson, in *Classical Electrodynamics* (Wiley, ed. 3, 1999), chap. 14.
29. T. Wulle, S. Herminghaus, *Phys. Rev. Lett.* **70**, 1401–1404 (1993).
30. E. Greenfield, M. Segev, W. Walasik, O. Raz, *Phys. Rev. Lett.* **106**, 213902 (2011).
31. M. Henstridge *et al.*, *Optica* **5**, 678 (2018).
32. Materials, methods, and supplemenatry media are available online.
33. T. E. Stevens, J. K. Wahlstrand, J. Kuhl, R. Merlin, *Science* **291**, 627–630 (2001).
34. B. M. Bolotovskiĭ, V. P. Bykov, *Sov. Phys. Usp.* (Engl. transl.) **33**, 477 (1990). [translation from *Usp. Fiziol. Nauk* **160**, 141 (1990)].
35. H. Ardavan, A. Ardavan, J. Singleton, *J. Opt. Soc. Am. A* **20**, 2137–2155 (2003).

ACKNOWLEDGMENTS

Funding: This work was supported by the MRSEC Program of the National Science Foundation under grant DMR-1120923 and the Air Force Office of Scientific Research under MURI grant FA9550-14-1-0389. The $LiTaO_3$ substrates were prepared at the Lurie Nanofabrication Facility, University of Michigan, and the metasurface was fabricated at the Birck Nanotechnology Center, Purdue University. **Author contributions:** R.M. conceived and

supervised the project; M.H. designed and carried out the ultrafast optical measurements, performed the calculations, and prepared the $LiTaO_3$ substrates; C.P. designed the metasurface with input from A.G.; D.W. fabricated the metasurface with contributions from V.M.S. and A.B.; and M.H. wrote the manuscript with help from R.M. All authors reviewed and commented on the manuscript. **Competing interests:** We declare no competing interests. **Data and materials availability:** All data are presented in the manuscript and supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6413/439/suppl/DC1
Materials and Methods
Supplementary Text
Fig. S1
Movie S1
References (*36*–*38*)

14 March 2018; accepted 30 August 2018
10.1126/science.aat5915

GAS SEPARATION

Ethane/ethylene separation in a metal-organic framework with iron-peroxo sites

Libo Li[1,2,6]*, Rui-Biao Lin[2]*, Rajamani Krishna[3], Hao Li[2,4], Shengchang Xiang[4], Hui Wu[5], Jinping Li[1,6]†, Wei Zhou[5]†, Banglin Chen[2]†

The separation of ethane from its corresponding ethylene is an important, challenging, and energy-intensive process in the chemical industry. Here we report a microporous metal-organic framework, iron(III) peroxide 2,5-dioxido-1,4-benzenedicarboxylate [$Fe_2(O_2)$(dobdc) (dobdc^{4-}: 2,5-dioxido-1,4-benzenedicarboxylate)], with iron (Fe)–peroxo sites for the preferential binding of ethane over ethylene and thus highly selective separation of C_2H_6/C_2H_4. Neutron powder diffraction studies and theoretical calculations demonstrate the key role of Fe-peroxo sites for the recognition of ethane. The high performance of $Fe_2(O_2)$(dobdc) for the ethane/ethylene separation has been validated by gas sorption isotherms, ideal adsorbed solution theory calculations, and simulated and experimental breakthrough curves. Through a fixed-bed column packed with this porous material, polymer-grade ethylene (99.99% pure) can be straightforwardly produced from ethane/ethylene mixtures during the first adsorption cycle, demonstrating the potential of $Fe_2(O_2)$(dobdc) for this important industrial separation with a low energy cost under ambient conditions.

Ethylene (C_2H_4) is the largest feedstock in petrochemical industries, with a global production capacity of more than 170 million tons in 2016. It is usually produced by steam cracking or thermal decomposition of ethane (C_2H_6), in which a certain amount of C_2H_6 residue coexists in the product and needs to be removed to produce polymer-grade (≥99.95% pure) C_2H_4 as the starting chemical for many other products, particularly the widely utilized polyethylene. The well-established industrial separation technology of the cryogenic high-pressure distillation process is one of the most energy-intensive processes in the chemical industry, requiring large distillation columns with 120 to 180 trays and high reflux ratios because of the similar sizes and volatilities of C_2H_4 and C_2H_6 (*1*, *2*). Realization of cost- and energy-efficient C_2H_4/C_2H_6 separation to obtain polymer-grade C_2H_4 is highly desired and has been recently highlighted as one of the most important industrial separation tasks for future energy-efficient separation processes (*3*–*5*).

Adsorbent-based gas separation, through pressure swing adsorption (PSA), temperature swing adsorption, or membranes, is a promising technology to replace the traditional cryogenic distillation and thus to fulfill the energy-efficient separation economy. Some adsorbents, such as γ-Al_2O_3 (*6*), zeolite (*7*, *8*), and metal-organic frameworks (MOFs) (*9*, *10*), have been developed for C_2H_4/C_2H_6 adsorptive separation. These porous materials take up larger amounts of C_2H_4 than of C_2H_6, mainly because of the stronger interactions of the immobilized metal sites, such as Ag(I) and Fe(II), on the pore surfaces with unsaturated C_2H_4 molecules (*9*, *11*). Although these kinds of adsorbents exhibit excellent adsorption separation performance toward C_2H_4/C_2H_6 mixtures, with the selectivity up to 48.7 (*12*), production of high-grade C_2H_4 is still quite energy intensive. This is because C_2H_4, as the preferentially adsorbed gas, needs to be further desorbed to get the C_2H_4 product. To remove the unadsorbed and contaminated C_2H_6, at least four adsorption-desorption cycles through inert gas or a vacuum pump are necessary to achieve the purity limit required (≥99.95%) for the C_2H_4 polymerization reactor (*13*).

If C_2H_6 is preferentially adsorbed, the desired C_2H_4 product can be directly recovered in the adsorption cycle. Compared with C_2H_4-selective adsorbents, this approach can save approximately 40% of energy consumption (0.4 to 0.6 GJ/ton of ethylene) (*14*, *15*) on PSA technology for the C_2H_4/C_2H_6 separation. Although porous materials have been well established for gas separation and purification (*16*–*22*), those exhibiting the preferred C_2H_6 adsorption over C_2H_4 are scarce. To date, only a few porous materials for selective C_2H_6/C_2H_4 separation have been reported (*2*, *13*, *23*, *24*), with quite low separation selectivity and productivity.

To target MOFs with the preferential binding of C_2H_6 over C_2H_4, it is necessary to immobilize some specific sites for the stronger interactions with C_2H_6. Inspired by natural metalloenzymes and synthetic compounds for alkane C–H activation in which M-peroxo, M-hydroperoxo, and M-oxo [M = Cu(II), Co(III), and Fe (III/IV)] are active catalytic intermediates (*25*–*27*), we hypothesized that similar functional sites within MOFs might have stronger binding with alkanes than alkenes and thus could be utilized for the selective separation of C_2H_6/C_2H_4. In this regard, $Fe_2(O_2)$(dobdc), developed by Bloch *et al.* and containing iron(III)-peroxo sites on the pore surfaces, might be of special interest (*28*, *29*). We thus synthesized the $Fe_2(O_2)$(dobdc), studied its binding for C_2H_6,

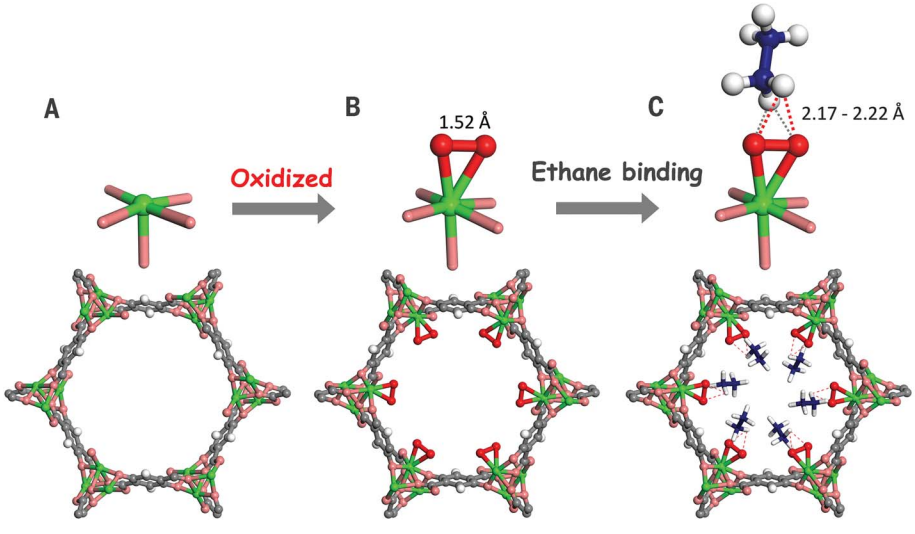


Fig. 1. Structures determined from NPD studies. Shown are structures of (**A**) Fe_2(dobdc), (**B**) $Fe_2(O_2)$(dobdc), and (**C**) $Fe_2(O_2)$(dobdc)⊃C_2D_6 at 7 K. Note the change from the open Fe(II) site to the Fe(III)-peroxo site for the preferential binding of ethane. Fe, green; C, dark gray; O, pink; O_2^{2-}, red; H or D, white; C in C_2D_6, blue.

[1]College of Chemistry and Chemical Engineering, Taiyuan University of Technology, Taiyuan 030024, Shanxi, China. [2]Department of Chemistry, University of Texas at San Antonio, One UTSA Circle, San Antonio, TX 78249-0698, USA. [3]Van 't Hoff Institute for Molecular Sciences, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, Netherlands. [4]Fujian Provincial Key Laboratory of Polymer Materials, College of Materials Science and Engineering, Fujian Normal University, Fuzhou 350007, Fujian, China. [5]NIST Center for Neutron Research, National Institute of Standards and Technology, Gaithersburg, MD 20899-6102, USA. [6]Shanxi Key Laboratory of Gas Energy Efficient and Clean Utilization, Taiyuan 030024, Shanxi, China.
*These authors contributed equally to this work.
†Corresponding author. Email: jpli211@hotmail.com (J.L.); wzhou@nist.gov (W.Z.); banglin.chen@utsa.edu (B.C.)

and examined the separation performance for C_2H_6/C_2H_4 mixtures. We found that $Fe_2(O_2)$(dobdc) exhibits preferential binding of C_2H_6 over C_2H_4. $Fe_2(O_2)$(dobdc) not only takes up moderately high amounts of C_2H_6 but also displays the highest C_2H_6/C_2H_4 separation selectivities in the wide pressure range among the examined porous materials, demonstrating it as the best material reported to date for this important gas separation to produce polymer-grade ethylene (99.99% pure).

$Fe_2(O_2)$(dobdc) was prepared according to the previously reported procedure with a slight modification (*28*). Both Fe_2(dobdc) and $Fe_2(O_2)$(dobdc) are air sensitive and need to be handled and stored in a dry box under an N_2 atmosphere. As expected, $Fe_2(O_2)$(dobdc) maintains the framework structure of Fe_2(dobdc) (Fig. 1, A and B, and fig. S1A), with a Brunauer-Emmett-Teller surface area of 1073 m^2/g (fig. S1B).

The C_2H_6 binding affinity in $Fe_2(O_2)$(dobdc) was first investigated by single-component sorption isotherms at a temperature of 298 K and pressures up to 1 bar, as shown in Fig. 2A. The C_2H_6 adsorption capacity on $Fe_2(O_2)$(dobdc) is much higher than that of C_2H_4, implying the distinct binding affinity of $Fe_2(O_2)$(dobdc) for C_2H_6. At 1 bar, the uptake amount of C_2H_6 in $Fe_2(O_2)$(dobdc) is 74.3 cm^3/g, corresponding to ~1.1 C_2H_6 per $Fe_2(O_2)$(dobdc) formula. Unlike the pristine Fe_2(dobdc), which takes up more C_2H_4 than C_2H_6 because of the Fe(II) open sites, $Fe_2(O_2)$(dobdc) adsorbs a larger amount of C_2H_6 than of C_2H_4. Therefore, we successfully realized the "reversed C_2H_6/C_2H_4 adsorption" in $Fe_2(O_2)$(dobdc) (fig. S2). The adsorption heats (Q_{st}) of C_2H_6 and C_2H_4 on $Fe_2(O_2)$(dobdc) were calculated by using the virial equation (fig. S3). The C_2H_6 adsorption heat of $Fe_2(O_2)$(dobdc) was calculated to be 66.8 kJ/mol at zero coverage, a much higher value than those reported for other MOFs (*2*), indicating the strong interaction between $Fe_2(O_2)$(dobdc) and C_2H_6 molecules. All of the isotherms are completely reversible and exhibit no hysteresis. Further adsorption cycling tests at 298 K (fig. S4) indicated no loss of C_2 uptake capacity over 20 adsorption-desorption cycles.

To structurally elucidate how C_2H_6 and C_2H_4 are adsorbed in this MOF, high-resolution neutron powder diffraction (NPD) measurements were carried out on C_2D_6-loaded and C_2D_4-loaded samples of $Fe_2(O_2)$(dobdc) at 7 K (see supplementary materials and fig. S5). As shown in Fig. 1C, C_2D_6 molecules exhibit preferential binding with the peroxo sites through C–D···O hydrogen bonds (D···O, ~2.17 to 2.22 Å). The D···O distance is much shorter than the sum of van der Waals radii of oxygen (1.52 Å) and hydrogen (1.20 Å) atoms, indicating a relatively strong interaction, which is consistent with the high C_2H_6 adsorption heat found in $Fe_2(O_2)$(dobdc). In addition, we noticed that, sterically, the nonplanar C_2D_6 molecule happens to match better to the uneven pore surface in $Fe_2(O_2)$(dobdc) than the planar C_2D_4 molecule (fig. S6), resulting in stronger hydrogen bonds with the Fe-peroxo active site and stronger van der Waals interactions with the ligand surface. To further understand the mechanism of the selective C_2H_6/C_2H_4 adsorption in $Fe_2(O_2)$(dobdc), we conducted detailed first-principles dispersion-corrected density functional theory calculations (see supplementary materials and table S1). The optimized C_2H_6 binding configuration on the Fe-peroxo site agrees reasonably well with the C_2D_6-loaded structures determined from the NPD data, indicating that the reversed C_2H_6/C_2H_4 adsorption selectivity originates from the peroxo active sites and the electronegative surface oxygen distribution in $Fe_2(O_2)$(dobdc). Similar preferential binding of C_2H_6 over C_2H_4 has also been experimentally found in another oxidized MOF, Cr-BTC(O_2) (where BTC is 1,3,5-benzenetricarboxylate) (figs. S7 and S8) (*30*).

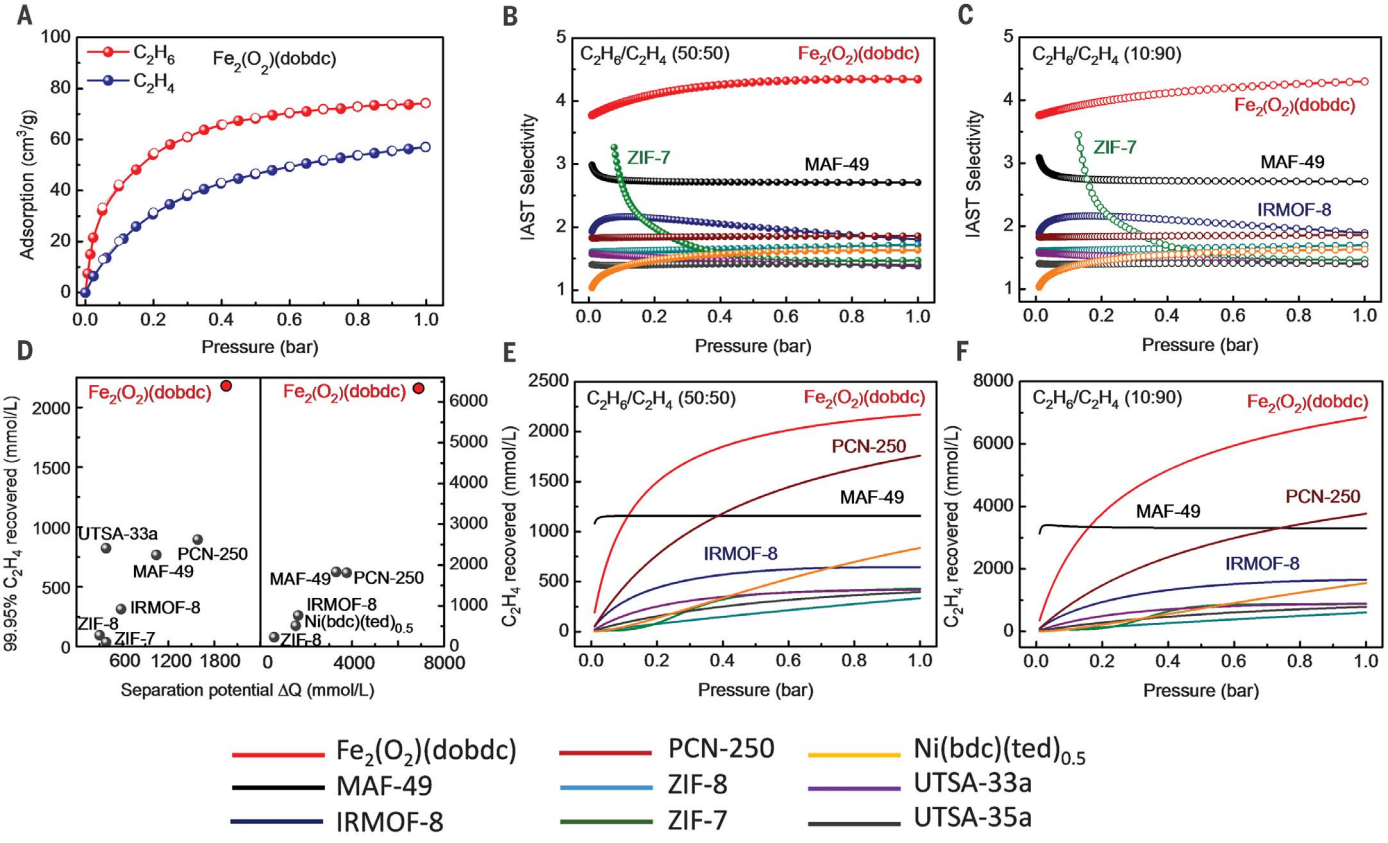


Fig. 2. C_2H_6 and C_2H_4 adsorption isotherms of $Fe_2(O_2)$(dobdc), IAST calculations, and separation potential simulations on C_2H_6-selective MOFs. (**A**) Adsorption (solid) and desorption (open) isotherms of C_2H_6 (red circles) and C_2H_4 (blue circles) in $Fe_2(O_2)$(dobdc) at 298 K. (**B** and **C**) Comparison of the IAST selectivities of $Fe_2(O_2)$(dobdc) with those of previously reported best-performing materials for C_2H_6/C_2H_4 (50/50 and 10/90) mixtures. (**D**) Predicted productivity of 99.95% pure C_2H_4 from C_2H_6/C_2H_4 (50/50 and 10/90) mixtures in fixed-bed adsorbers at 298 K. (**E** and **F**) Separation potential of $Fe_2(O_2)$(dobdc) for C_2H_6/C_2H_4 [50/50 (E) and 10/90 (F)] mixtures versus those of best-performing MOFs.

Ideal adsorbed solution theory (IAST) calculations were performed to estimate the adsorption selectivities of C_2H_6/C_2H_4 (50/50 and 10/90) for $Fe_2(O_2)$(dobdc) and other C_2H_6-selective materials (Fig. 2B). The fitting details are provided in the supplementary materials (figs. S9 to S17 and tables S2 to S11). Compared with other top-performing MOFs [MAF-49, IRMOF-8, ZIF-8, ZIF-7, PCN-250, Ni(bdc)(ted)$_{0.5}$, UTSA-33a, and UTSA-35a], $Fe_2(O_2)$(dobdc) exhibits a new benchmark for C_2H_6/C_2H_4 (50/50) adsorption selectivity (4.4) at 1 bar and 298 K, greater than the selectivity of the previously reported best-performing MOF, MAF-49 (2.7) (*2*). This value is also higher than the highest value (2.9) among 30,000 all-silica zeolite structures that were investigated by Kim *et al.* through computational screening (*31*). For a C_2H_6/C_2H_4 (10/90) mixture, under the same conditions, $Fe_2(O_2)$(dobdc) also exhibits the highest adsorption selectivity among these MOFs (Fig. 2C).

Next, transient breakthrough simulations were conducted to validate the feasibility of using $Fe_2(O_2)$(dobdc) in a fixed bed for separation of C_2H_6/C_2H_4 mixtures (fig. S18). Two C_2H_6/C_2H_4 mixtures (50/50 and 10/90) were used as feeds to mimic the industrial process conditions. The simulated breakthrough curves show that C_2H_6/C_2H_4 (50/50) mixtures were completely separated by $Fe_2(O_2)$(dobdc), whereby C_2H_4 breakthrough occurred first within seconds to yield the polymer-grade gas and then C_2H_6 passed through the fixed bed after a certain time (τ_{break}). To evaluate the C_2H_6/C_2H_4 separation ability of these MOFs, the separation potential ΔQ was calculated to quantify the mixture separations in fixed-bed adsorbers (table S12). Attributed to the record-high C_2H_6/C_2H_4 selectivity and relatively high C_2H_6 uptake, the amount of 99.95% pure C_2H_4 recovered by $Fe_2(O_2)$(dobdc) reached up to 2172 mmol/liter (C_2H_6/C_2H_4, 50/50) and 6855 mmol/liter (C_2H_6/C_2H_4, 10/90) (Fig. 2D), values which are almost two times higher than those for the other benchmark materials. $Fe_2(O_2)$(dobdc) has the highest separation potential for recovering the pure C_2H_4 from (50/50) C_2H_6/C_2H_4 mixtures during the adsorption process (Fig. 2E). Even when the concentration of C_2H_6 decreases to 10% (Fig. 2F), $Fe_2(O_2)$(dobdc) maintains the highest separation potential (table S13), which makes it the most promising material for the separation of C_2H_6 from C_2H_6/C_2H_4 mixtures.

These excellent breakthrough results from simulation encouraged us to further evaluate the separation performance of $Fe_2(O_2)$(dobdc) in the actual separation process. Several breakthrough experiments were performed on an in-house–constructed apparatus, which was described in our previous work (*32*). The breakthrough experiments were performed on several selected MOFs, including $Fe_2(O_2)$(dobdc), with C_2H_6/C_2H_4 (50/50) mixtures flowed over a packed bed at a total flow rate of 5 ml/min at 298 K (fig. S19 and table S14). For $Fe_2(O_2)$(dobdc), a clean and sharp separation of C_2H_6/C_2H_4 was observed (Fig. 3A). C_2H_4 was first to elute through the bed, before it was contaminated with undetectable amounts of C_2H_6, resulting in a high concentration of C_2H_4 feed that was ≥99.99% pure (the detection limit of the instrument is 0.01%). After some period, the adsorbent got saturated, C_2H_6 broke through, and then the outlet gas stream quickly reached equimolar concentrations. To make the systematic comparison for the C_2H_4 separation performance in the selected MOFs, C_2H_4 purity and productivity were calculated from their breakthrough curves (table S15). For $Fe_2(O_2)$(dobdc), 0.79 mmol/g of C_2H_4 with ≥99.99% purity can be recovered from the C_2H_4/C_2H_6 (50/50) mixture in a single breakthrough operation; this value is nearly three times that for the benchmark material MAF-49 (0.28 mmol/g). In addition, the cycle and regeneration capabilities of $Fe_2(O_2)$(dobdc) were further studied by breakthrough cycle experiments (Fig. 3B), with no noticeable decrease in the mean residence times for both C_2H_6 and C_2H_4 within five continuous cycles under ambient conditions. Moreover, $Fe_2(O_2)$(dobdc) material retained its stability after the breakthrough cycling test (fig. S20).

In the real production of high-purity C_2H_4, the C_2H_6 concentration in C_2H_4/C_2H_6 mixtures produced by naphtha cracking is about 6 to 10%, and the feed gases are also contaminated by low levels of impurities such as CH_4, H_2, and C_2H_2 (*33*). Therefore, breakthrough experiments on C_2H_6/C_2H_4 (10/90) mixtures and $C_2H_6/C_2H_4/CH_4/H_2/C_2H_2$ (10/87/1/1/1) mixtures were also performed for $Fe_2(O_2)$(dobdc). As shown in Fig. 3, C and D, highly efficient separations for both mixtures were realized, which further demonstrates that $Fe_2(O_2)$(dobdc) can be used to purify C_2H_4 with low concentrations of C_2H_6 even in the presence of CH_4, H_2, and C_2H_2 impurities.

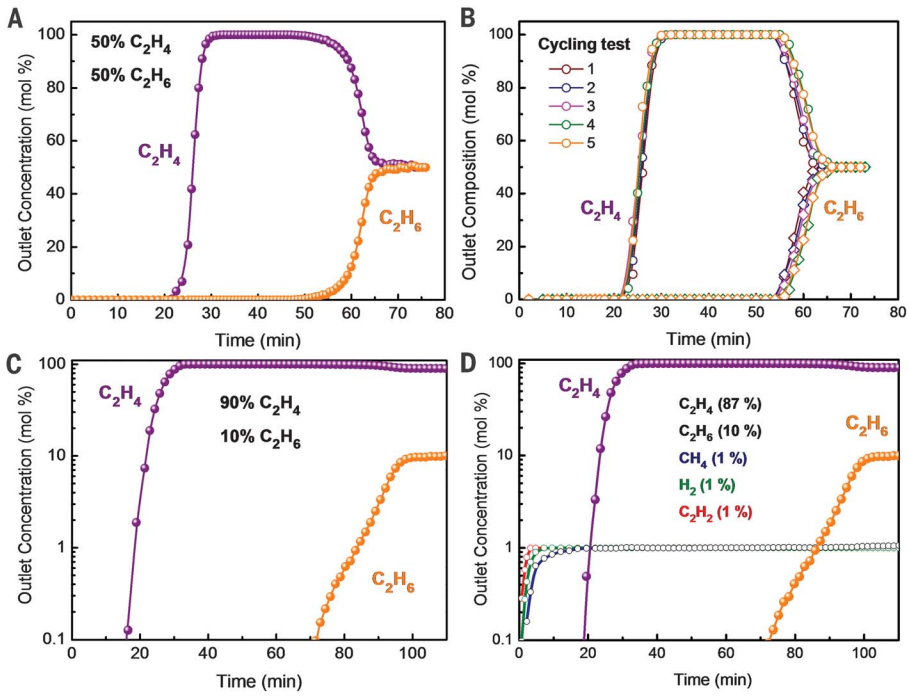


Fig. 3. Breakthrough experiments. Experimental column breakthrough curves for (**A**) a C_2H_6/C_2H_4 (50/50) mixture, (**B**) a cycling test of C_2H_6/C_2H_4 (50/50) mixtures, (**C**) C_2H_6/C_2H_4 (10/90) mixtures, and (**D**) $C_2H_6/C_2H_4/C_2H_2/CH_4/H_2$ (10/87/1/1/1) mixtures in an absorber bed packed with $Fe_2(O_2)$(dobdc) at 298 K and 1.01 bar.

In summary, we discovered that a distinctive MOF with Fe-peroxo sites can induce stronger interactions with C_2H_6 than with C_2H_4, leading to the unusual reversed C_2H_6/C_2H_4 adsorption. The fundamental binding mechanism of $Fe_2(O_2)$(dobdc) for the recognition of C_2H_6 has been demonstrated through neutron diffraction studies and theoretical calculations, indicating the important role of the Fe-peroxo sites for the preferential interactions with C_2H_6. This material can readily produce high-purity C_2H_4 (≥99.99% pure) from C_2H_4/C_2H_6 mixtures during the first breakthrough cycle with moderately high productivity and a low energy cost. The strategy we developed in this work may be broadly applicable, which will facilitate extensive research on the immobilization of different sites into porous MOFs for stronger interactions with C_2H_6 than with C_2H_4, thus targeting some practically useful porous materials with low material costs and high productivity for the practical industrial realization of this very challenging and important separation.

REFERENCES AND NOTES

1. F. Bander, Separation Technologies; http://seperationtechnology.com.
2. P.-Q. Liao, W.-X. Zhang, J.-P. Zhang, X.-M. Chen, *Nat. Commun.* **6**, 8697 (2015).
3. D. S. Sholl, R. P. Lively, *Nature* **532**, 435–437 (2016).
4. J. Y. S. Lin, *Science* **353**, 121–122 (2016).
5. S. Chu, Y. Cui, N. Liu, *Nat. Mater.* **16**, 16–22 (2016).

6. R. T. Yang, E. S. Kikkinides, *AIChE J.* **41**, 509–517 (1995).
7. P. J. Bereciartua *et al*., *Science* **358**, 1068–1071 (2017).
8. S. Aguado, G. Bergeret, C. Daniel, D. Farrusseng, *J. Am. Chem. Soc.* **134**, 14635–14637 (2012).
9. E. D. Bloch *et al*., *Science* **335**, 1606–1610 (2012).
10. Y. He, R. Krishna, B. Chen, *Energy Environ. Sci.* **5**, 9107–9120 (2012).
11. B. Li *et al*., *J. Am. Chem. Soc.* **136**, 8654–8660 (2014).
12. S. Yang *et al*., *Nat. Chem.* **7**, 121–129 (2015).
13. C. Gücüyener, J. van den Bergh, J. Gascon, F. Kapteijn, *J. Am. Chem. Soc.* **132**, 17704–17706 (2010).
14. A. Mersmann, B. Fill, R. Hartmann, S. Maurer, *Chem. Eng. Technol.* **23**, 937–944 (2000).
15. T. Ren, M. Patel, K. Blok, *Energy* **31**, 425–451 (2006).
16. H. Furukawa, K. E. Cordova, M. O'Keeffe, O. M. Yaghi, *Science* **341**, 1230444 (2013).
17. H. Sato *et al*., *Science* **343**, 167–170 (2014).
18. A. Cadiau, K. Adil, P. M. Bhatt, Y. Belmabkhout, M. Eddaoudi, *Science* **353**, 137–140 (2016).
19. R. Vaidhyanathan *et al*., *Science* **330**, 650–653 (2010).
20. P. Nugent *et al*., *Nature* **495**, 80–84 (2013).
21. Q.-G. Zhai *et al*., *Nat. Commun.* **7**, 13645 (2016).
22. K. Li *et al*., *J. Am. Chem. Soc.* **131**, 10368–10369 (2009).
23. W. Liang *et al*., *Chem. Eng. Sci.* **148**, 275–281 (2016).
24. Y. Chen *et al*., *Chem. Eng. Sci.* **175**, 110–117 (2018).
25. D. J. Xiao *et al*., *Nat. Chem.* **6**, 590–595 (2014).
26. Z. Li *et al*., *J. Am. Chem. Soc.* **139**, 15251–15258 (2017).
27. J. Cho *et al*., *Nature* **478**, 502–505 (2011).
28. E. D. Bloch *et al*., *J. Am. Chem. Soc.* **133**, 14814–14822 (2011).
29. W. L. Queen *et al*., *Dalton Trans.* **41**, 4180–4187 (2012).
30. L. J. Murray *et al*., *J. Am. Chem. Soc.* **132**, 7856–7857 (2010).
31. J. Kim *et al*., *Langmuir* **28**, 11914–11919 (2012).
32. L. Li *et al*., *J. Am. Chem. Soc.* **139**, 7733–7736 (2017).
33. R. A. Meyers, *Handbook of Petrochemicals Production Processes* (McGraw-Hill, 2005).

ACKNOWLEDGMENTS

L.L. and R.-B.L. thank B. Li for the discussions on this project and S. Li for preparation of breakthrough experiments. **Funding:** We gratefully acknowledge the financial support from the National Natural Science Foundation of China (21606163), the Natural Science Foundation of Shanxi (201601D021042), and the Welch Foundation (AX-1730). **Author contributions:** L.L., R.-B.L., J.L., W.Z., and B.C. conceived the idea and designed the experiments. L.L. synthesized the materials and carried out most of the adsorption and separation experiments. H.L. and S.X. prepared the samples and analyzed the data. R.K. calculated the IAST selectivity and performed the simulated breakthrough. L.L., W.Z., and H.W. carried out the NPD experiments and analyzed the results. L.L., R.-B.L., W.Z., and B.C. interpreted the results and wrote the paper. **Competing interests:** None declared. **Data and materials availability:** Crystallographic data reported in this paper are provided in the supplementary materials and archived at the Cambridge Crystallographic Data Centre under reference numbers 1817715 to 1817716, 1574716 to 1574717, and 1859806 to 1859808. All other data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6413/443/suppl/DC1
Materials and Methods
Figs. S1 to S20
Tables S1 to S15
References (*34*–*40*)

19 January 2018; resubmitted 8 June 2018
Accepted 5 September 2018
10.1126/science.aat0586

QUANTUM OPTICS

A room-temperature single-photon source based on strongly interacting Rydberg atoms

Fabian Ripka, Harald Kübler, Robert Löw, Tilman Pfau*

Tailored quantum states of light can be created via a transfer of collective quantum states of matter to light modes. Such collective quantum states emerge in interacting many-body systems if thermal fluctuations are overcome by sufficient interaction strengths. Therefore, ultracold temperatures or strong confinement are typically required. We show that the exaggerated interactions between Rydberg atoms allow for collective quantum states even above room temperature. The emerging Rydberg interactions lead both to suppression of multiple Rydberg state excitations and destructive interference due to polariton dephasing. We experimentally implemented a four-wave mixing scheme to demonstrate an on-demand single-photon source. The combination of glass cell technology, identical atoms, and operation around room temperature promises scalability and integrability. This approach has the potential for various applications in quantum information processing and communication.

Single-photon emitters are of interest for manifold applications in quantum computation, simulation, sensing, and especially in quantum secure communication. In particular, the latter will demand efficient quantum repeaters (*1*) that require efficient single-photon creation and storage schemes, ideally at the same platform. Room-temperature alkali gases can perfectly serve as such storage media, and second-long storage times have already been demonstrated for weak coherent states (*2*). However, the generation of single photons with stable wavelength and adapted bandwidth is still a challenge.

The first observation of antibunched photon counting statistics from single-atom fluorescence dates back to 40 years ago (*3*). Today, very efficient on-demand solid-state sources based on quantum dots (*4*), color centers (*5*), and single molecules (*6*) are available. Although in principle most of these systems are working at room temperature as well (*7*–*9*), they do require cryogenic temperatures for optimal performance. Because they are solid-state embedded, they suffer from interactions with phonons, spin noise, strain, and drifting electric fields. These result in large variations in frequencies, spectral wandering, and additional phase noise causing spectral diffusion, which is a fundamental limitation for scalability.

In contrast, atoms (*3*) and ions (*10*) inherently produce spectrally indistinguishable single photons. Here, the highest fidelities have been achieved with ultracold atoms inside high-finesse cavities (*11*). Small photonic networks have been realized with them (*12*). With room-temperature gases, so far, heralded single photons with sub-Doppler linewidth can be created (*13*). There, two photons are generated in a probabilistic process; one of them then gives the information on the existence of the other one. However, as with parametric down-conversion (*14*), they undergo a spontaneous generation process, and this requires a triggered memory for temporal synchronization of the created photons.

A rather new approach makes use of strongly interacting Rydberg atoms, which suppress multiple excitations within a certain radius (*15*). Related to this, Rydberg interactions can also lead to dephasing between Rydberg polaritons (*16*, *17*). These interaction mechanisms can be converted into an effective optical nonlinearity, even on the single-photon level (*18*). On the basis of the Rydberg blockade effect in ultracold samples, researchers have demonstrated photon antibunching

5. Physikalisches Institut, Universität Stuttgart, Center for Integrated Quantum Science and Technology, 70569 Stuttgart, Germany.
***Corresponding author. Email: t.pfau@physik.uni-stuttgart.de**

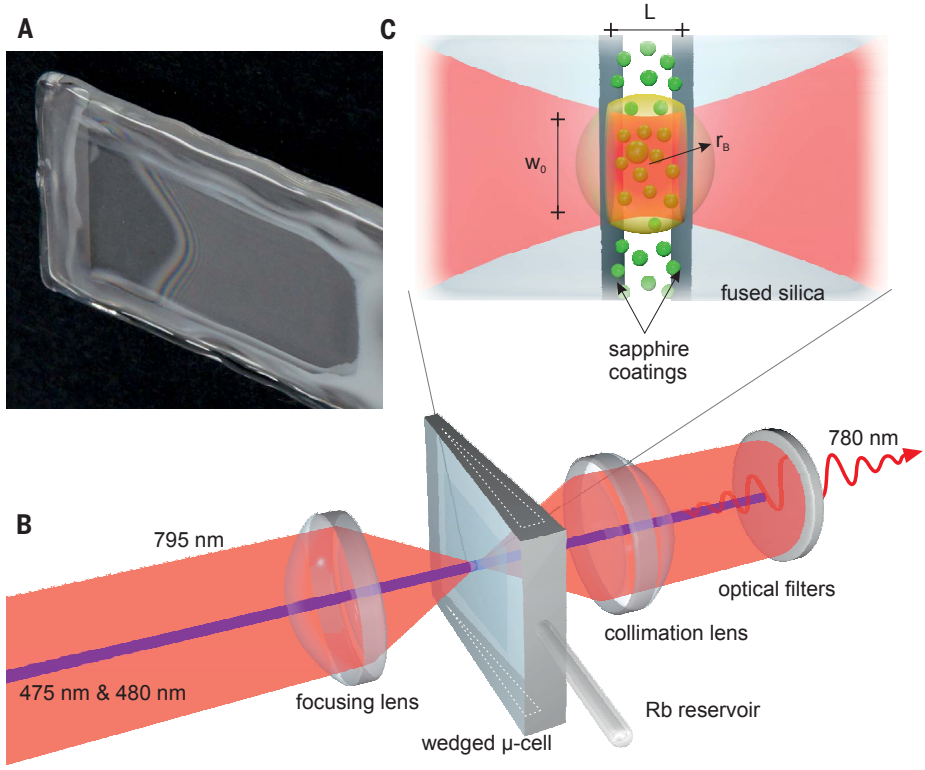


Fig. 1. Experimental setup. (**A**) Photograph of the wedge-shaped microcell. The outer dimensions are 25 mm × 50 mm. The glass windows have a thickness of 1.21 ± 0.01 mm, matching the high–numeric aperture aspheric lenses (focusing and collimation), so that their diffraction limits are reached. The inner distance of the glass windows varies from touching each other to about 100 μm. The thicknesses are determined via residual-finesse resonances. In the region of 400 nm to few μm, interference effects can be observed (so-called Newton's rings). (**B**) Schematic of the setup. (**C**) The excitation volume inside the vapor cell. The excitation of the rubidium atoms is transversally limited to the diameter of the 795-nm beam; 475- and 480-nm beams are much larger; all optical beams are copropagating (*32*).

(*19*, *20*) and even higher-order correlated light states (*21*).

However, Rydberg-excited atoms can be coherently excited and optically probed not only at ultralow temperatures but also in thermal vapors (*22*). Nanosecond-pulsed and intense excitation fields at GHz Rabi coupling strengths allow for coherent dynamics faster than 1 ns per Rabi cycle (*23*). On this time scale, thermal atoms move less than the optical wavelength and can be assumed to be nearly frozen. As soon as they are excited, the collective excitation can be retrieved within 1.2 ns, before dephasing due to spatial inhomogeneities and the Doppler effect set in (*24*). For strong Rydberg interaction effects to be observable, the atom-atom interaction strength must be larger than the excitation bandwidth. The equivalent of 3 GHz of van der Waals–type Rydberg interactions can be achieved in a dephasing regime where the excitation volume is larger than the interaction distance (*25*).

Blockade of Rydberg excitations occurs at distances shorter than the blockade radius $r_B = (C_6/\hbar\Omega_{eff})^{1/6}$, where C_6 is the van der Waals coefficient, $\hbar$ is the reduced Planck constant, and Ω_{eff} is the effective two-photon coupling strength (*26*). If we assume that the addressed Rydberg state is $40S_{1/2}$ and that $\Omega_{eff} = 400$ MHz, then (*27*) $C_6 = 2\pi\hbar \times 806.3$ MHz μm^6 and the blockade radius results in $r_B = 1.1 \pm 0.1$ μm. For multiple excitations in an atomic ensemble to be suppressed by Rydberg blockade, the size of the ensemble must be smaller than that radius. Rubidium atoms can be coherently excited to Rydberg states in vapor cells of such a size, without being limited by atom-wall interactions (*28*). Therefore, it has been predicted that a four-wave mixing experiment in a microscopic cell will result in a suppression of multiple excitations and hence to an emission of single photons (*29*). Note that even when multiple Rydberg excitations cannot be prevented, the superposition of the shifted excitations leads to destructive interference in the detection mode (*16*). It is also important to note that the size of such a small ensemble is still larger than the optical wavelength. This guarantees a directed emission pattern during the four-wave mixing process (*30*). Besides, the emitted photon reflects the shapes of the excitation pulses and is not created in a spontaneous emission process. Furthermore, the emitted photons do not undergo any jitter in either the center frequency or the bandwidth.

To truncate the Rydberg-excited ensembles to μm^3 volumes, our setup focuses the 795-nm excitation beam to a waist with $w_0 = 1.45 \pm 0.05$ μm Gaussian $1/e^2$ diameter (Fig. 1C). The vapor cell is designed in a wedge-shaped geometry and provides a distance range from 0 to several tens of micrometers (Fig. 1, A and B). Thus, we can choose the degree of longitudinal confinement of the atomic ensemble. The cell is filled with rubidium at natural abundance and the experiment is done using the ^{85}Rb isotope. The vapor pressure is set by a temperature of 140°C, which corresponds to a steady-state density of approximately 30 atoms per μm^3 in the hyperfine ground state $F = 3$. Yet this is not sufficient for an efficient single-photon generation. However, because a large number of atoms are attached to the surfaces of the glass windows, we desorb a fraction of them via light-induced atomic desorption (*31*). A 2.5-ns pulse at 480 nm raises the number of atoms to more than 1000 per μm^3 within 5 ns (Fig. 2B). The Rydberg-excited volume then contains about 2000 atoms. In the following few nanoseconds, the optical thickness decreases again as they hit the opposite surface.

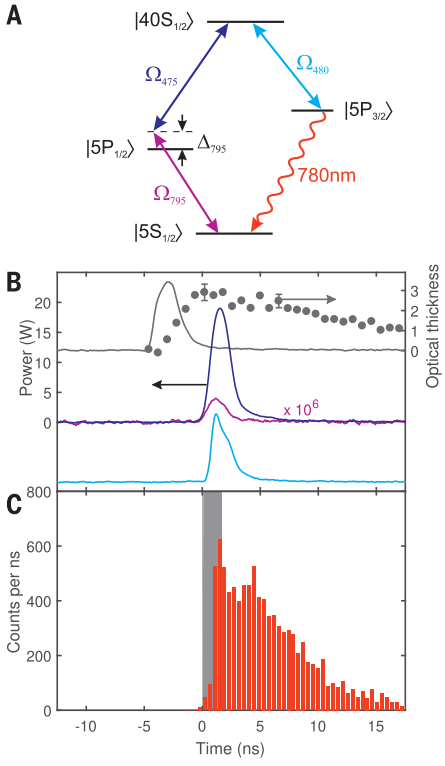


Fig. 2. Experiment scheme. (**A**) Excitation level scheme. The excitation path is clockwise starting from $5P_{1/2}$ to $5P_{3/2}$ via $40S_{1/2}$. (**B**) Pulse sequence. The desorption pulse (gray) leads to a drastic increase of the optical depth within 5 ns (gray dots; error bars denote SD). The optical thickness is determined via statistical measurements at the single-photon level. Magenta and blue, excitation pulses; cyan, deexcitation pulse. Desorption pulse and deexcitation pulse are plotted with an offset relative to the power axis (left). (**C**) Typical signal shape retrieved from a cell length of 1.14 μm. The bin width is 0.425 ns. Nonclassical emission is observed in an early time interval of the signal, marked in gray (*32*).

The photons at 780 nm are generated in a pulsed four-wave mixing cycle about 5 ns after the pulsed desorption (Fig. 2, A and B). The durations of the excitation pulses are 2.5 ns each. In the limit of adiabatic elimination of the intermediate $5P_{1/2}$ state, the Rabi frequency of the coupling between the ground state and the Rydberg state is $\Omega_{eff}/2\pi = 0.41 \pm 0.16$ GHz. The Rabi frequency of the transition from the Rydberg state to the excited state $5P_{3/2}$ is $\Omega_{480}/2\pi = 1.15 \pm 0.09$ GHz. The photons at 780 nm are emitted into the same spatial mode as the 795-nm beam. They are collected by an aspheric high–numeric aperture lens behind the cell. For spatial and spectral filtering, the emission is coupled into a single-mode fiber and frequency-filtered by bandpass filters and a temperature-stabilized etalon with a bandwidth of 1.7 GHz. Afterward, the photons are detected in a Hanbury Brown and Twiss–type setup. The excitation pulses are recorded in order to enable a post-selection of events with respect to intensities and timing jitters (*32*).

Figure 2C shows a histogram of photoelectric detection events. The signal is mainly composed of two kinds of origins: (i) early photons that are coherently generated in the four-wave mixing process with a decay time of around 1 ns, and (ii) partly incoherently generated photons that take over at larger delays after the excitation pulses and are mainly originating from collisions of excited atoms with the glass windows. Both processes are much faster than the natural lifetime of the excited atoms. These two kinds of photons differ not only in their temporal appearance but also in their spatial emission pattern. The reason is that the four-wave mixing photons emit into the excitation mode, whereas the background photons are assumed to emit in the full solid angle.

In Fig. 3A, a normalized second-order intensity correlation function $g^{(2)}(\tau)$ is shown, where the discrete delay is given in multiples of the repetition time (20 ms). The detection events are limited to a time window of 0 to 1.7 ns (gray area in Fig. 2C). All measurements from cell thicknesses 0.94 to 1.14 μm are concatenated. At zero delay, we observe clear antibunching with $g^{(2)}(\tau = 0) = 0.21^{+10}_{-5}$, where the error is the statistical uncertainty given by the error propagation of ±1 standard deviation of the correlation fluctuations at $\tau \neq 0$. The four-wave mixing should give $g^{(2)} = 1$, so this constitutes clear evidence for strong Rydberg interactions that lead to excitation blockade and polariton dephasing. The detection probability in the measurement shown in Fig. 3A is on average 0.60%. Accounting for all attenuating elements between generation (behind the first lens) and detection of the photons, such as the coupling into a single-mode fiber, imperfect spectral filtering, and finite detection efficiencies of the single-photon counters, we obtain a mean generation efficiency (brightness) of $\varepsilon = 4.0\%$.

The effect of the excitation volume can be seen in Fig. 3B. Here, the zero-delay photon-pair correlation is depicted as a function of the cell length L. The blue data points show pair correlations in the time window 0 to 1.7 ns; the red data points show the time span of 1.7 to 10 ns as a reference. When the cell length is below a characteristic length $L_0 = 1.16 \pm 0.01$ μm, Rydberg-Rydberg interaction will prevent the emission of more than one photon into the detection mode. This value L_0 is related to but not equivalent to the blockade radius. As a comparison to the

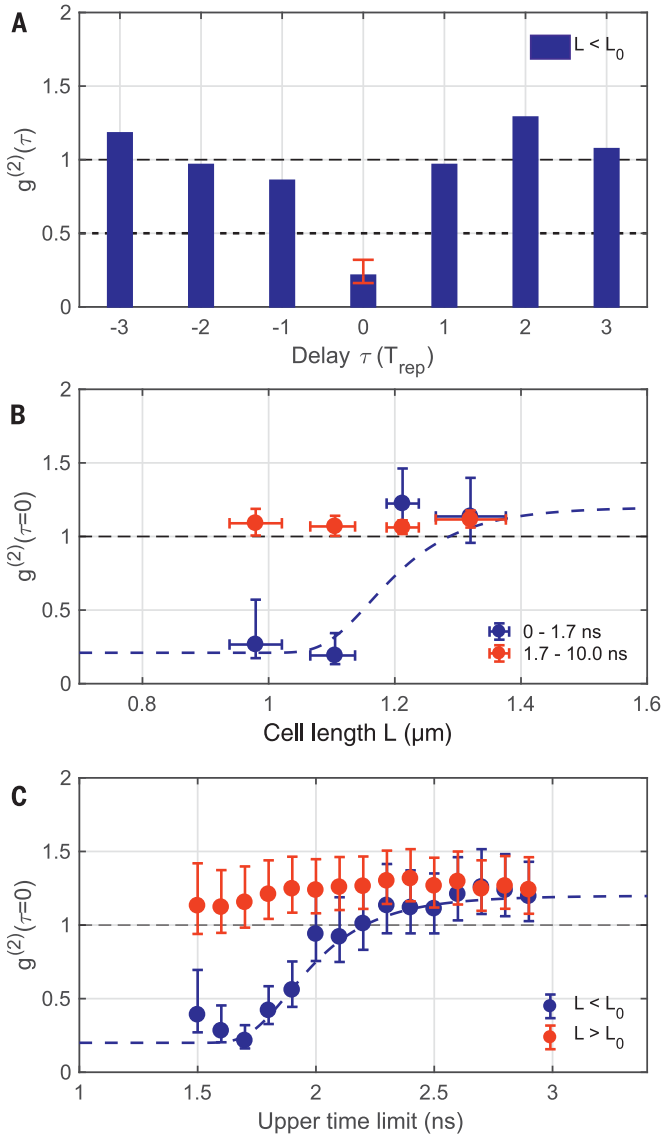


Fig. 3. Normalized second-order photon correlation functions. (**A**) Normalized second-order intensity correlation function $g^{(2)}(\tau)$ as a function of coincidence delay time in multiples of the cycle repetition time (20 ms); time window, 0 to 1.7 ns; cell thickness $L < L_0$; $g^{(2)}(\tau = 0) = 0.21^{+10}_{-5}$. (**B**) Dependence on the cell thickness in two time windows: 0 to 1.7 ns (blue), and 1.7 to 10 ns (red) as a reference. The blue dashed line is a guide to the eye based on a generic model $g^{(2)}(0) = (1 - g^{(2)}{}_{bg}) \exp[-(L_0/L)^{\alpha}] + g^{(2)}{}_{bg}$ (*19*), which is related to the hard-sphere character of the blockade. The characteristic length is $L_0 = 1.16$ μm. (**C**) Dependence on the upper limit of the time window, for cell thicknesses $L < L_0$ (blue) and $L > L_0$ (red) as a reference. The blue dashed line is a guide to the eye and is described by a generic model similar to that of Fig. 3B. Error bars in the *y* direction are obtained by error propagation of ±1 SD due to statistical fluctuations; error bars in the *x* direction mark the span of *x*-positions of the cell setting the individual data points.

strong interactions, we also investigated the $32S_{1/2}$ Rydberg state with $r_B = 0.85 \pm 0.13$ μm. We obtained Poissonian light statistics for all cell lengths around 1 μm and above.

The evolution of the imprint of the Rydberg interactions with time can be seen in Fig. 3C. It shows the photon-pair correlation versus the upper limit of the correlation time window that has a constant duration of 1.7 ns. At cell lengths smaller than L_0 (blue), $g^{(2)}(\tau = 0)$ reveals antibunching again. The coherent many-body state lasts about 1.9 ns, after which the imprint of the antibunching vanishes. As a comparison with negligible Rydberg interaction, the photon-pair correlation at larger cell lengths $L > L_0$ is shown in red.

We have shown that strong Rydberg interactions in a hot atomic vapor cell lead to the generation of single photons. This has two consequences: (i) Strong atomic interaction effects do not need ultracold temperatures to be observable in the laboratory, and (ii) a single-photon emitter can be realized with promising features for future applications. Further experimental work will be needed to determine the indistinguishability of the generated photons, apply optical integration technologies, and combine this source with photon memories based on atomic vapors.

REFERENCES AND NOTES

1. L. M. Duan, M. D. Lukin, J. I. Cirac, P. Zoller, *Nature* **414**, 413–418 (2001).
2. O. Katz, O. Firstenberg, *Nat. Commun.* **9**, 2074 (2018).
3. H. J. Kimble, M. Dagenais, L. Mandel, *Phys. Rev. Lett.* **39**, 691–695 (1977).
4. P. Michler *et al.*, *Science* **290**, 2282–2285 (2000).
5. C. Kurtsiefer, S. Mayer, P. Zarda, H. Weinfurter, *Phys. Rev. Lett.* **85**, 290–293 (2000).
6. C. Brunel, B. Lounis, P. Tamarat, M. Orrit, *Phys. Rev. Lett.* **83**, 2722–2725 (1999).
7. A. Beveratos *et al.*, *Eur. Phys. J. D* **18**, 191–196 (2002).
8. P. Michler *et al.*, *Nature* **406**, 968–970 (2000).
9. B. Lounis, W. E. Moerner, *Nature* **407**, 491–493 (2000).
10. F. Diedrich, H. Walther, *Phys. Rev. Lett.* **58**, 203–206 (1987).
11. A. Kuhn, M. Hennrich, G. Rempe, *Phys. Rev. Lett.* **89**, 067901 (2002).
12. S. Ritter *et al.*, *Nature* **484**, 195–200 (2012).
13. C. Shu *et al.*, *Nat. Commun.* **7**, 12783 (2016).
14. P. G. Kwiat *et al.*, *Phys. Rev. Lett.* **75**, 4337–4341 (1995).
15. M. D. Lukin *et al.*, *Phys. Rev. Lett.* **87**, 037901 (2001).
16. F. Bariani, P. M. Goldbart, T. A. B. Kennedy, *Phys. Rev. A* **86**, 041802(R) (2012).
17. H. Busche *et al.*, *Nat. Phys.* **13**, 655–658 (2017).
18. O. Firstenberg, C. S. Adams, S. Hofferberth, *J. Phys. B* **49**, 152003 (2016).
19. Y. O. Dudin, A. Kuzmich, *Science* **336**, 887–889 (2012).
20. T. Peyronel *et al.*, *Nature* **488**, 57–60 (2012).
21. Q.-Y. Liang *et al.*, *Science* **359**, 783–786 (2018).
22. A. K. Mohapatra, T. R. Jackson, C. S. Adams, *Phys. Rev. Lett.* **98**, 113003 (2007).
23. B. Huber *et al.*, *Phys. Rev. Lett.* **107**, 243001 (2011).
24. F. Ripka, Y.-H. Chen, R. Löw, T. Pfau, *Phys. Rev. A* **93**, 053429 (2016).
25. T. Baluktsian, B. Huber, R. Löw, T. Pfau, *Phys. Rev. Lett.* **110**, 123001 (2013).
26. R. Löw *et al.*, *J. Phys. B* **45**, 113001 (2012).
27. S. Weber *et al.*, *J. Phys. B* **50**, 133001 (2017).
28. H. Kübler, J. P. Shaffer, T. Baluktsian, R. Löw, T. Pfau, *Nat. Photonics* **4**, 112–116 (2010).
29. M. M. Müller *et al.*, *Phys. Rev. A* **87**, 053412 (2013).
30. M. Saffman, T. G. Walker, *Phys. Rev. A* **66**, 065403 (2002).
31. S. N. Atutov *et al.*, *Phys. Rev. A* **67**, 053401 (2003).
32. See supplementary materials.

ACKNOWLEDGMENTS

We thank Y.-H. Chen for her contributions in the early stage of the experiment, F. Schreiber for crafting the vapor cell, N. Gruhler for the coating of the vapor cell with sapphire, C. Müller and M. Gryzlova for contributions to Fig. 1, P. Michler and M. Schwartz for useful discussions about single-photon sources, and H. Alaeian for reading and commenting the manuscript. **Funding:** Supported by the ERC under contract 267100 and the BMBF within Q.com-Q (project 16KIS0129). **Author contributions:** The experiment was conceived by all authors; execution of the experiment as well as data analysis were performed by F.R.; F.R. wrote the manuscript with contributions from all authors. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6413/446/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 and S2
References (*33*–*35*)

16 May 2018; accepted 30 August 2018
10.1126/science.aau1949

SOLAR CELLS

Methylammonium-free, high-performance, and stable perovskite solar cells on a planar architecture

Silver-Hamill Turren-Cruz[1,2,3], Anders Hagfeldt[2]*, Michael Saliba[1]*

Currently, perovskite solar cells (PSCs) with high performances greater than 20% contain bromine (Br), causing a suboptimal bandgap, and the thermally unstable methylammonium (MA) molecule. Avoiding Br and especially MA can therefore result in more optimal bandgaps and stable perovskites. We show that inorganic cation tuning, using rubidium and cesium, enables highly crystalline formamidinium-based perovskites without Br or MA. On a conventional, planar device architecture, using polymeric interlayers at the electron- and hole-transporting interface, we demonstrate an efficiency of 20.35% (stabilized), one of the highest for MA-free perovskites, with a drastically improved stability reached without the stabilizing influence of mesoporous interlayers. The perovskite is not heated beyond 100°C. Going MA-free is a new direction for perovskites that are inherently stable and compatible with tandems or flexible substrates, which are the main routes commercializing PSCs.

Perovskites with an ABX_3 structure [A = cesium (Cs), methylammonium (MA), formamidinium (FA); B = lead (Pb), tin (Sn); X = chlorine (Cl), bromine (Br) , iodine (I)] have skyrocketed in recent years, with power conversion efficiencies (PCEs) now at 23.3%, which is close to established technologies such as gallium arsenide (GaAs), cadmium telluride (CdTe), and silicon (Si) (*1*). Currently, almost all high-performance perovskite solar cells (PSCs) with efficiencies >20%, including all reported world records, contain unstable MA (*2*–*6*).

One important trend to achieve PCEs >20% has been the usage of more complex perovskite compositions, ranging from double-cation (MAFA or CsFA) (*7*–*9*) and triple-cation (CsMAFA) (*6*) to Rb-modified perovskites (*5*, *10*–*12*). FA is used as the majority cation in almost all current high-efficiency PSCs because FA, compared with MA, is thermally more stable (*8*, *13*) and has a more optimal, red-shifted bandgap because it is the largest organic cation to still fit into a three-dimensional perovskite (*14*). Unfortunately, the relatively large size of FA also induces a distorted lattice that results in a photoinactive "yellow phase" at room temperature, whereas the photoactive "black phase" can only be observed at elevated temperatures (*8*). Therefore, obtaining phase-stable, black-phase FA perovskites at room temperature has become the implicit or explicit objective of the perovskite research field. Avoiding yellow-phase impurities motivated the development of the above-mentioned optimized processing techniques and elaborate multication, multihalide mixtures. The addition of Br has been especially crucial to suppress yellow-phase impurities. Accordingly, Br often constitutes up to 20% of the precursor compositions within most high-performance PSCs (*5*, *6*, *15*, *16*).

However, using the halide position to improve crystallization comes at a cost: Br blueshifts the bandgap disproportionately. We illustrate this point in Fig. 1A, where we show the Shockley-Queisser limit for the theoretical, maximum PCE and short-circuit current (J_{SC}) as a function of the bandgap (E_g) (*17*). By interchanging Br with I in $MAPbI_xBr_{1-x}$ compounds, a bandgap range from 1.58 eV ($MAPbI_3$) to 2.28 eV ($MAPbBr_3$), spanning 700 meV, is achieved (*18*).

In stark contrast, the cation interacts only moderately with the metal-halide cage and accordingly has a relatively small bandgap shift of only 100 meV from $MAPbI_3$ (1.58 eV) to $FAPbI_3$ (1.48 eV) (*18*). The instability of $MAPbI_3$ is reported by various works that show film degradation because of degassing MA (*13*, *19*–*22*). In Fig. 1B, we illustrate the volatile nature of the MA molecule itself using the reported degradation pathway of CH_3NH_3I into CH_3I and NH_3

[1]Adolphe Merkle Institute, Chemin des Verdiers 4, CH-1700 Fribourg, Switzerland. [2]Laboratory of Photomolecular Science, Institute of Chemical Sciences and Engineering, École Polytechnique Fédérale de Lausanne, Station 6, CH-1015 Lausanne, Switzerland. [3]Benemérita Universidad Autónoma de Puebla. CIDS, Av. San Claudio y 18 Sur, Col. San Manuel, Ciudad Universitaria, CP 72570, P.O. Box 1067, Puebla, Pue. 7200, Mexico.
*Corresponding author. Email: anders.hagfeldt@epfl.ch (A.H.); miliba@gmail.com (M.S.)

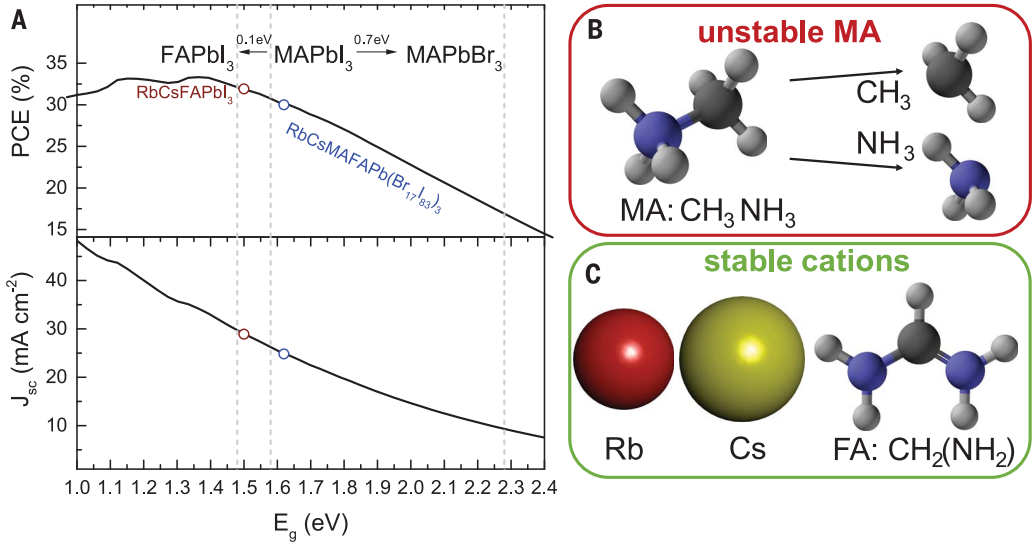


Fig. 1. Disproportionate bandgap tuning and cation stability. (**A**) Shockley-Queisser limit for the theoretical, maximum PCE and short-circuit current (J_{SC}) as a function of the bandgap (E_g). A wide bandgap range is spanned from 1.48 eV ($FAPbI_3$) to 1.58 eV ($MAPbI_3$) to 2.28 eV ($MAPbBr_3$). The cation position (FA to MA) shifts the bandgap only moderately by 0.1 eV, whereas the halide position (I to Br) incurs a disproportionate "blue penalty," shifting the bandgap by 0.7 eV. The red open circle is the newly developed, non-MA, non-Br compound $RbCsFAPbI_3$, which is closer to more optimal bandgaps, whereas $RbCsMAFAPb(Br_{17}I_{83})_3$ (blue open circle) (*5*), one of the currently highest-performing compositions, has a suboptimal bandgap and contains unstable MA. (**B**) Illustration of the volatile nature of the MA molecule. One reported degradation pathway is CH_3NH_3I into CH_3I and NH_3 (the I atom is omitted for simplicity) at temperatures as low as 80°C, as reported in (*23*). (**C**) By contrast, Rb, Cs, and FA are thermally more stable cations and therefore preferable for long-term stability.

at temperatures as low as 80°C (the I atom is omitted in the figure for simplicity) (*23*). Therefore, MA remains a principled risk factor for long-term stability and should thus be avoided (although the exact time scales of MA degradation require further research). To substantiate this point further, supplementary text 1 (together with figs. S1 to S4) shows the XRD and respective ultraviolet–visible (UV-vis) spectra before and after annealing at 130°C for 3 hours of MA-containing perovskite films—$MAPbI_3$, $Cs_5MAFAPbI_3$, $MAFAPbI_3$, and $Rb_5Cs_5MAFAPbI_3$—and non-MA perovskite films—$FAPbI_3$, Rb_5FAPbI_3, $Cs_{10}FAPbI_3$, and $Rb_5Cs_{10}FAPbI_3$. All MA-based perovskites exhibit strong degradation, as evidenced by the growth of the PbI_2 peak at 12°, a behavior which does not occur for the non-MA perovskites. Although thin films are not full devices, they nevertheless provide a clear motivation to avoid MA in all future compositions, ensuring perovskites with intrinsic long-term stability that is required for decade-long stable solar cells. Industry cannot afford the long-term risk factor of MA. However, currently almost all reported PCEs >20%, including all those setting world records, use MA (*2*–*6*), whereas perovskites without MA currently have PCEs often below 20% and frequently use high-temperature steps (*2*, *5*, *24*).

Thus, a more ideal perovskite compound ought to avoid Br because of the "blue penalty" and MA for stability reasons. This leaves I as the preferred halide and the thermally more stable Rb, Cs, and FA as the preferred cations (Fig. 1C), rendering $RbCsFAPbI_3$ perovskites particularly relevant. Hence, in this work, we explored $Rb_xCs_yFA_{(100-x-y)}PbI_3$ compositions. We discovered an optimized $RbCsFAPbI_3$ perovskite (without MA and Br) with a bandgap of 1.53 eV, close to the single-junction optimum, with high short-circuit currents comparing favorably with the previously used $RbCsMAFAPb(Br_{17}I_{83})_3$ perovskite (Fig. 1A, red and blue open circles). We achieved a planar PSC with a short-circuit current of 25.06 mA cm^{-2} and a high PCE of 20.44% (stabilized at 20.35%) for planar PSCs, which is among the highest non-MA–containing results thus far reported. The perovskites do not require any heating beyond 100°C, rendering them compatible with perovskite/Si tandem solar cells or flexible solar cells, which are among the most attractive pathways to commercialization.

We provide x-ray diffraction (XRD), photoluminescence (PL), and UV-vis absorption data for $Rb_xCs_yFA_{(100-x-y)}PbI_3$ perovskite films (on a planar substrate), written for convenience as Rb_xCs_yFAI going forward (x and y are in percentages throughout). This nomenclature refers to the precursor solutions as outlined in the supplementary materials.

We started our investigation with the double-cation compounds $Cs_yFA_{(100-y)}I$ and $Rb_xFA_{(100-x)}I$. In supplementary text 2 (together with figs. S5 to S8), the XRD data for $Cs_yFA_{(100-y)}I$ (figs. S5 and S6) show the characteristic perovskite peak at 14° that shifts as more Cs is added. This corresponds to a blue-shift in the PL and UV-vis data as shown in Fig. 2A and as previously reported (*8*, *25*). For $Rb_xFA_{(100-x)}I$, the perovskite peak in the XRD data does not shift as strongly as for the Cs series (figs. S7 and S8). This is consistent with the PL and UV-vis data (Fig. 2B), in which only a very slight blue-shift occurs for small Rb quantities (Fig. 2B). As the amount of Rb is increased, only a very small blue-shift in the PL occurs. Albeit for $Rb_{15}FA_{85}I$, a more noticeable blue-shift and broadening occur in the PL, correlating with the occurrence of a second phase in the XRD at 10° (fig. S7). In addition, in contrast to CsFA, the $RbFAPbI_3$ films require annealing at 150°C, close to the temperature needed to convert pure $FAPbI_3$ films. This is consistent with previous work (*5*, *11*), confirming that there is no "black phase" for $RbPbI_3$.

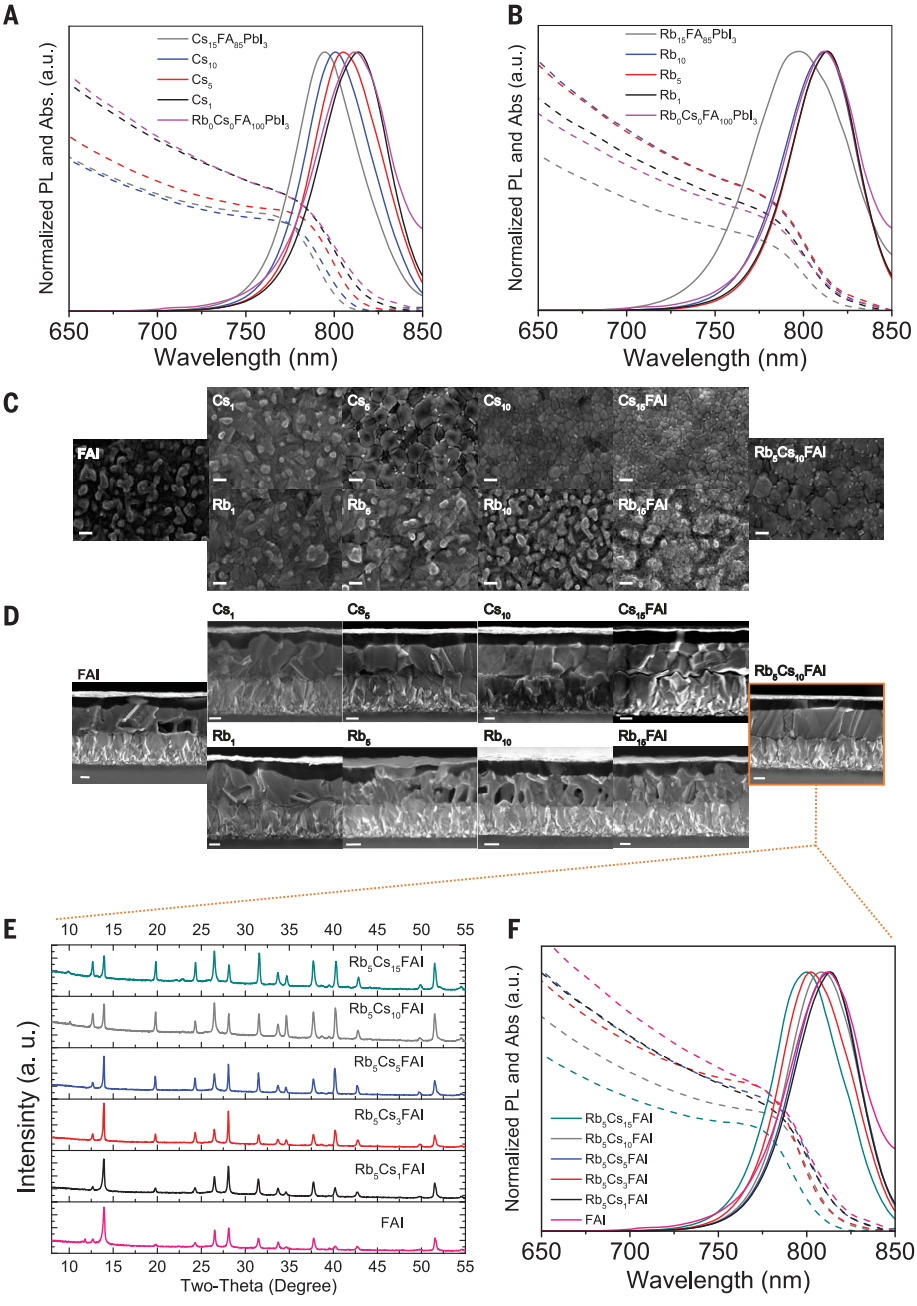


Fig. 2. Inorganic cation tuning: UV-vis, PL, SEM, and XRD characterization. (**A** and **B**) PL and UV-vis data for the (A) $Cs_yFA_{(100-y)}PbI_3$ (x = 0, 1, 5, 10, or 15%) and the (B) $Rb_xFA_{(100-x)}PbI_3$ (y = 0, 1, 5, 10, or 15%) series. (**C**) Corresponding top-view SEM images. Scale bar, 200 nm. (**D**) Corresponding cross-sectional SEM images for full devices. Scale bar, 200 nm. The last image on the right is the $Rb_5Cs_{10}PbI_3$ composition that results later in the highest device results. (**E**) XRD data and (**F**) PL and UV-vis data for the $Rb_xCs_yFA_{(100-x-y)}PbI_3$ series.

From scanning electron microscopy (SEM) top-view images in Fig. 2C, we observed that $Cs_yFA_{(100-y)}I$ films, starting with 5% Cs, have a relatively ordered film morphology with large grains. By contrast, Rb, likely because of the lacking integration, has a less ordered film morphology, resembling pure $FAPbI_3$ films. In Fig. 2C, far right, $Rb_5Cs_{10}FAI$, has a regular film morphology. We display specifically $Rb_5Cs_{10}FAI$ because it shows the highest and most reproducible device results.

We fabricated full devices on a planar stack of glass/FTO/SnO_2/perovskite/spiro/Au as reported previously (*26*). In Fig. 2D, cross-sectional SEM images are presented. The pure $FAPbI_3$ film is highly irregular and rough. For the CsFA series, starting with $Cs_5FA_{95}PbI_3$, the grains become monolithic throughout the film, reaching from the top to the bottom contact. For the RbFA series, for all Rb concentrations, the films resemble $FAPbI_3$, which is not beneficial for the perovskite/hole transporter materials (HTM) interface. The $Rb_5Cs_{10}FAI$ compound has particularly monolithic, large grains.

Last, in Fig. 2, E and F, we present XRD, PL, and UV-vis data for the $Rb_5Cs_yFA_{(95-y)}I$ series. The most blue-shifted composition is for 15% Cs followed, surprisingly, by 3% Cs. Upon adding 5 and 10% Cs, the PL remains relatively close to that of pure FA-based perovskite. This is consistent with Rb and Cs interacting with one another. We posit Cs could be aiding Rb to modify the lattice more effectively, but it is also possible that for certain Rb/Cs ratios, $RbCsPbI_3$ perovskites form (in Fig. 2E, shown is the occurrence of an additional phase at 10° for 10 and 15% Cs). These inorganic compounds could "lock up" excess I-halides and help form a passivation layer for the remaining, pure $FAPbI_3$ perovskite that has a more optimal single-junction bandgap. This is most noteworthy for the $Rb_5Cs_{10}FAI$ compound, which has a PL emission very close to $FAPbI_3$, in contrast to the non–Rb-containing $Cs_{10}FAI$ (Fig. 2A).

From device data in supplementary text 3 (together with figs. S9 to S16), we observed relatively high-performing CsFA (figs. S9 and S11) compounds that are not matched by the analogous RbFA series (figs. S10 and S12).

The double-cation CsFA films, possibly because of the large mismatch between the small Cs and large FA, do not yield highly reproducible PSCs. Although CsFA can occasionally reach high PCEs of 19.23% (fig. S9), we show in fig. S14A that the statistical baseline is lower (*5*). The Rb-modification again yields higher-performing perovskites with PCEs toward 19.52% for $Rb_5Cs_{10}FAI$ (the RbCsFAI series is provided in fig. S13). In fig. S14, we show that the best-performing $Cs_{10}FAI$ perovskite improves upon 5% Rb addition in terms of absolute efficiencies and especially reproducibility (an often neglected metric).

From supplementary text 3, we identified $Rb_5Cs_{10}FAPbI_3$ as a promising composition (figs. S13 and S14) and therefore continued with this composition for the remainder of this study. We successfully red-shifted the bandgap as motivated in Fig. 1. This can be observed clearly in the external quantum efficiency (EQE) measurements in figs. S15 and S16, where the EQE follows the PL trends. Consequently, we measured a high short-circuit current density of 24.52 mA cm^{-2}, as confirmed with EQE data that show an integrated J_{SC} of 24.01 mA cm^{-2} (fig. S16B), highlighting the inorganic cation tuning as a versatile tool to achieve more optimized, MA-free bandgaps.

The planar device architecture was used because it is inherently more compatible with flexible and tandem applications (none of the processing steps exceeds 100°C). Currently, planar PSCs show relatively high PCEs >20% and up to 21.4% (all using MA) (*27*–*29*), which is slightly lower than mesoporous-containing perovskites at 22.1% (also using unstable MA) (*4*). One reason for this is that planar PSCs are prone to exhibit shunting pathways between the highly monolithic perovskite grains (*30*). Accordingly, one of the main parameters lacking for planar PSCs is the fill factor (FF) (supplementary text 2). In planar PSCs, any pinhole in the perovskite film results in a shunt-path that enables a direct pathway between electron and hole transporter. By contrast, an additional meso-layer avoids direct contact between the electron and hole transporter, preventing pinholes from becoming detrimental shunt paths. Consequently, for planar PSCs, the perovskite layer itself needs to be formed nearly perfectly, and thus the processing window for planar PSCs is much narrower than

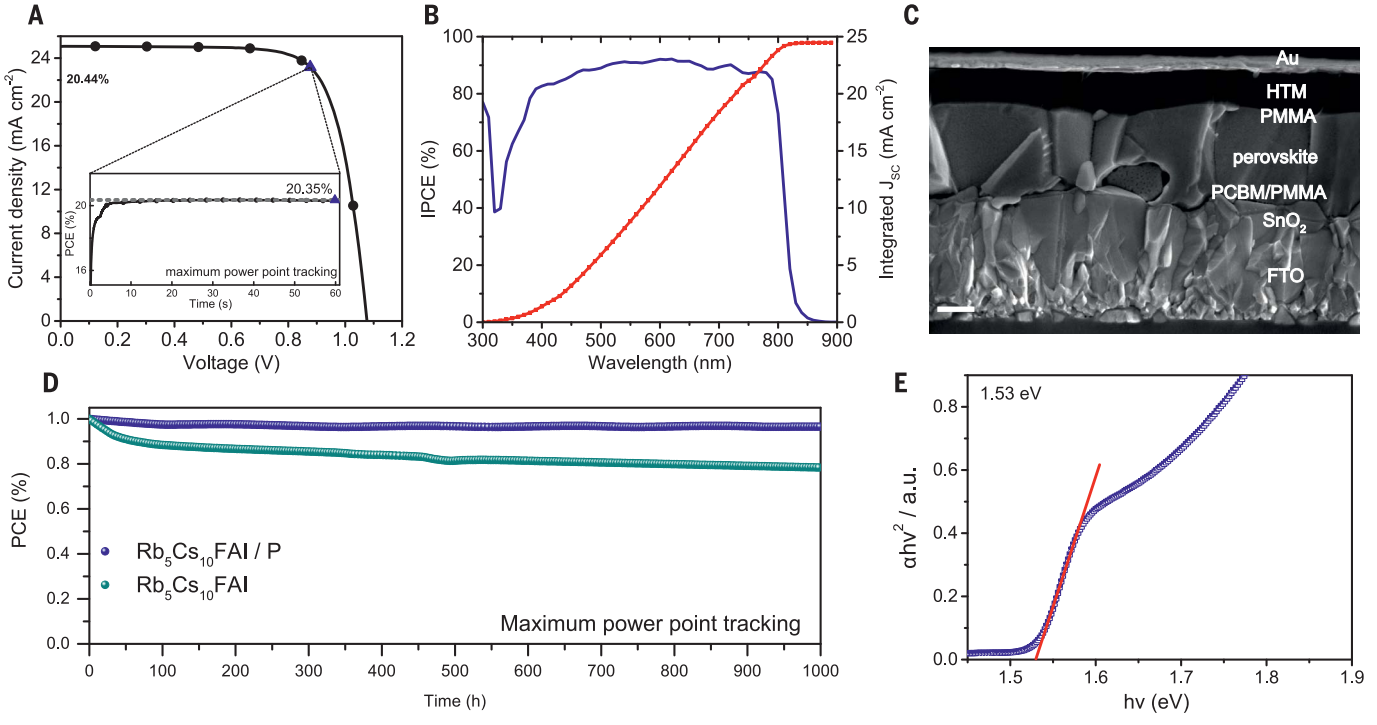


Fig. 3. Best-performing device, *J-V*, EQE, SEM, and stability. (**A**) *J-V* curve with MPP of a $Rb_5Cs_{10}FAPbI_3$ device with polymer layers composed of PCBM/PMMA at the SnO_2/perovskite interface and PMMA at the perovskite/HTM interface. The planar device displays a stabilized power output of 20.35%. (**B** and **C**) Corresponding (B) EQE and (C) SEM image. (**D**) Stability for the $Cs_{10}Rb_5FAPbI_3$ device without polymer layers (green curve) and with polymer-modification (blue curve) aged at room temperature after 1000 hours of continuous MPP tracking in a nitrogen atmosphere. (**E**) Tauc plot resulting in a 1.53-eV bandgap for the optimized $Rb_5Cs_{10}FAPbI_3$ perovskite.

for meso PSCs. In addition, we posit that the formation of pinholes is less likely with a meso-layer because of a more uniformly surface wettability for the liquid precursor in contrast to planar PSCs, for which surface wettability is a constant concern. Thus, an imperfect perovskite layer can be saved by the meso interlayer. In other words, meso PSCs are more forgiving of imperfections. Consequently, because of the challenging fabrication, planar PSCs have remained less efficient than meso PSCs. This results in fewer groups that specialize on planar PSCs, which in turn results in fewer reported results. At the moment, this is a vicious circle: Because planar PSCs do not reach world records yet, groups are less willing to switch from meso to planar, even though it is industrially more preferable (*31*). To change this, we aimed for high-performance, MA-free perovskites on a planar architecture, requiring modifications of the planar architecture.

Recently, much progress has been made in improving perovskite interfaces through polymeric buffer layers that help to passivate the perovskite, reduce recombination, and mitigate the negative impact from pinholes as well as block metal electrode migration at elevated temperatures (*5*, *32*). This approach is used both for the electron-transporting layer (ETL)/perovskite and the HTM/perovskite interface—for example, C60-SAM (*33*), phenyl-C61-butyric acid methyl ester (PCBM) (*34*), and poly(methyl methacrylate) (PMMA) (*32*). As another example, mixing PCBM and PMMA was shown to benefit the ETL/perovskite interface (*35*). This method, however, still uses an undesirable high-temperature titanium dioxide (TiO_2) mesoporous layer. Therefore, we explored the potential of polymeric buffer layers for a planar ETL/perovskite interface using our optimized $Rb_5Cs_{10}FAI$ compound (supplementary text 4 and figs. S17 to S19). In fig. S17, we explored different concentrations of PCBM at the SnO_2/perovskite interface, realizing that no improvement was achieved compared with the control device without polymeric layers. We then investigated different concentrations of a mixed polymeric interlayer of PCBM:PMMA at the SnO_2/perovskite interface (fig. S18). For planar PSCs, we found that PCBM:PMMA with a 5:1 ratio increases the average FF with values up to 77% (when compared with the control device without polymer-modification). The reproducibility, a very critical parameter for planar devices, is also improved, as indicated by the lowered standard deviation.

Encouraged by this, we investigated the top interface between the HTL and the perovskite using a PMMA layer on devices that already have a PCBM:PMMA polymer mix at the ETL interface. We show the results for the double-polymer–modified, planar PSCs in fig. S19. The additional PMMA layer improves the device architecture further. Perovskite films with and without a PMMA layer were characterized further by using atomic force microscopy (fig. S20) and contact angle measurements (fig. S21). From these measurements, we can conclude that the surface roughness is reduced and the contact angle increased when using PMMA, indicating that the polymer layer indeed modified the interface.

Whereas the devices with a 10:0.1 PCBM:PMMA show some of the highest PCE, the 5:1 PCBM:PMMA ratio exhibits more reproducible device parameters, especially in the FF. Thus, we used double-polymer–modified PSCs (pm-PSCs) with a 5:1 PCBM:PMAA mix at the ETL/perovskite and a PMMA layer between the perovskite/HTM as the most promising system for high performances and reproducibility.

In figs. S19 and S14B, we evaluated statistically the potential of the FTO/SnO_2/PCBM:PMMA/perovskite/PMMA/HTM stack. We observed in fig. S19 a similar average J_{SC} at 24.67 mA cm^{-2} (control) and 24.51 mA cm^{-2} (polymer-modified), a similar open-circuit voltage (V_{OC}) from 1095 to 1083 mV, and an increase of the average FF from 68 to 74%, translating into an improved average PCE from 18.20 to 19.71%. The polymer-modification improves reproducibility, which is especially noticeable in the FF.

In Fig. 3A, we show the champion device for pm-PSCs reaching a high stabilized power output of 20.35% [the current-voltage (*J-V*) parameters are provided in table S1]. The J_{SC} of 25.06 mA cm^{-2} is in good agreement with the integrated EQE value of 24.48 mA cm^{-2} (Fig. 3B). (In addition, figs. S22 and S23 display a 0.5 cm^2 area device and *J-V* parameters for differently sized masks, respectively). In Fig. 3C, we show the SEM image of a representative $Rb_5Cs_{10}FAPbI_3$ polymer-modified device. Moreover, in Fig. 3E, we show a Tauc plot with a bandgap of 1.53 eV for $Rb_5Cs_{10}FAPbI_3$, confirming once more that the inorganic cation tuning achieved the original goal of red-shifting the bandgap.

Stability is one of the most important obstacles remaining for PSCs to become attractive for industry. However, relatively few stability datasets exist for planar devices.

We therefore measured stability as outlined in previous works (*5*, *27*, *36*). We used the $Rb_5Cs_{10}FAPbI_3$ composition with and without the polymer modification, applying aging at room temperature for 1000 hours of continuous maximum power point (MPP) tracking in a nitrogen atmosphere. As shown in Fig. 3D, without polymeric buffer layers, the loss is relatively high (starting at 19.54% and finishing at 15.31% PCE). In stark contrast, the polymer-modified device has a only small loss. The absolute PCE starts at 18.74%, drops quickly to 17.63% after 35 hours (sometimes compared with a "burn-in"), and then exhibits a long-term component that stays relatively constant, finishing at 17.51% PCE after 1000 hours. Thus, after the initial component, the efficiency loss is less than 2 relative %, which is among the best-reported stability data yet for planar PSCs (without using MA). The stability of the $Rb_5Cs_{10}FAPbI_3$ compound is confirmed in figs. S24 and S25, where additional devices, including mesoporous (figs. S25 and S26), were aged under the same conditions for 500 hours. This stability data highlights the importance of the device architecture for the long-term stability of especially planar PSCs. This implies also that the mesoporous titania interlayer in the current world-record architectures not only contributes to higher performances but also higher stabilities. We posit that the higher stability is due to the meso-layer acting as a mechanic barrier against external degradation factors (such as moisture, vapors, and a diffusing metal electrode) as well as the challenges for planar PSCs to deposit a nearly perfect, compact perovskite absorber, which is needed for high performances but also to protect against the mentioned external degradation factors. Thus, stabilizing planar architectures, without the help from protective mesoporous interlayers, is one of the key challenges for perovskite research. Therefore, the contacts, such as polymeric buffer layers, will play an especially important role in fortifying planar PSCs further in the future, as clearly evidenced in Fig. 3D (*37*).

More generally, using only inorganic additives to phase-stabilize the thermally relatively stable FA perovskites is a compositional design strategy that may be used to stabilize intermediate bandgaps as well, which is highly attractive for tandem applications. The prospect of a perovskite/Si tandem especially has inspired much progress in the field. However, Si solar cells are stable over many decades, and therefore, any unstable perovskite component, such as MA, must be avoided if PSCs were to be combined with Si. Our work provides a direct pathway for circumventing many of the drawbacks so far observed in MA- and Br-containing perovskites. Future strategies for MA-free perovskites can also include inorganic cation tuning using combinations of K, Rb, and Cs with FA (*38*, *39*).

REFERENCES AND NOTES

1. www.nrel.gov/pv/assets/images/efficiency-chart.png.
2. W. S. Yang *et al.*, *Science* **348**, 1234–1237 (2015).
3. S. S. Shin *et al.*, *Science* **356**, 167–171 (2017).
4. W. S. Yang *et al.*, *Science* **356**, 1376–1379 (2017).
5. M. Saliba *et al.*, *Science* **354**, 206–209 (2016).
6. M. Saliba *et al.*, *Energy Environ. Sci.* **9**, 1989–1997 (2016).
7. N. Pellet *et al.*, *Angew. Chem. Int. Ed. Engl.* **53**, 3151–3157 (2014).
8. J. W. Lee *et al.*, *Adv. Energy Mater.* **5**, 1501310 (2015).
9. D. P. McMeekin *et al.*, *Science* **351**, 151–155 (2016).
10. M. Zhang *et al.*, *ACS Ener. Lett.* **2**, 438–444 (2017).
11. Y. H. Park *et al.*, *Adv. Funct. Mater.* **27**, 1605988 (2017).
12. T. Duong *et al.*, *Nano Energy* **30**, 330–340 (2016).
13. B. Conings *et al.*, *Adv. Energy Mater.* **5**, 1500477 (2015).
14. T. M. Koh *et al.*, *J. Phys. Chem. C* **118**, 16458–16462 (2014).
15. N. J. Jeon *et al.*, *Nature* **517**, 476–480 (2015).
16. D. Q. Bi *et al.*, *Nat. Energy* **1**, 16142 (2016).
17. W. Shockley, H. J. Queisser, *J. Appl. Phys.* **32**, 510–519 (1961).
18. J. H. Noh, S. H. Im, J. H. Heo, T. N. Mandal, S. I. Seok, *Nano Lett.* **13**, 1764–1769 (2013).
19. W. Tan, A. R. Bowring, A. C. Meng, M. D. McGehee, P. C. McIntyre, *ACS Appl. Mater. Interfaces* **10**, 5485–5491 (2018).
20. S. H. Wang, Y. Jiang, E. J. Juarez-Perez, L. K. Ono, Y. B. Qi, *Nat. Energy* **2**, 16195 (2017).
21. T. Zhang *et al.*, *J. Mater. Chem. A Mater. Energy Sustain.* **5**, 1103–1111 (2017).
22. Y. C. Zhao *et al.*, *J. Phys. Chem. C* **121**, 14517–14523 (2017).
23. E. J. Juarez-Perez, Z. Hawash, S. R. Raga, L. K. Ono, Y. B. Qi, *Energy Environ. Sci.* **9**, 3406–3410 (2016).
24. Z. Wang *et al.*, *Nat. Energy* **2**, 17135 (2017).
25. Z. Li *et al.*, *Chem. Mater.* **28**, 284–292 (2016).
26. E. H. Anaraki *et al.*, *Energy Environ. Sci.* **9**, 3128–3134 (2016).
27. H. Tan *et al.*, *Science* **355**, 722–726 (2017).
28. Q. Jiang *et al.*, *Nat. Energy* **2**, 16177 (2016).
29. Q. Jiang *et al.*, *Adv. Mater.* **29**, 1703852 (2017).
30. J. P. Correa-Baena *et al.*, *Energy Environ. Sci.* **10**, 710–727 (2017).

31. M. Saliba *et al.*, *Chem. Mater.* **30**, 4193–4201 (2018).
32. F. J. Wang *et al.*, *J. Phys. Chem. C* **121**, 1562–1568 (2017).
33. K. Wojciechowski *et al.*, *J. Phys. Chem. Lett.* **6**, 2399–2405 (2015).
34. J. Xu *et al.*, *Nat. Commun.* **6**, 7081 (2015).
35. J. Peng *et al.*, *Energy Environ. Sci.* **10**, 1792–1800 (2017).
36. M. Saliba, *Science* **359**, 388–389 (2018).
37. M. Lira-Cantú, *Nat. Energy* **2**, 17115 (2017).
38. M. Abdi-Jalebi *et al.*, *Nature* **555**, 497–501 (2018).
39. D. J. Kubicki *et al.*, *J. Am. Chem. Soc.* **139**, 14173–14180 (2017).

ACKNOWLEDGMENTS

S.-H.T.-C. and M.S. thank the Adolphe Merkle Foundation for support. S.-H.T.-C. thanks CONACYT-México for support. S.-H.T.-C. and M.S. thank the Adolphe Merkle Foundation for support. A.H. thanks the Swiss National Science Foundation for financial support with the project entitled "Fundamental studies of mesoscopic devices for solar energy conversion" with project number 200021_157135/1. The authors thank E. Ochoa Martinez and J. A. Borrego Perez for technical assistance. **Funding:** This work was partially funded by the Adolphe Merkle Foundation (M.S.). **Author contributions:** M.S. conceived, designed, and led the overall project; S.-H.T.-C. fabricated all samples and devices and conducted PL, UV-vis, contact angle, and long-term aging experiments on the perovskite films and devices; M.S. conducted SEM measurements; and S.-H.T.-C. and M.S. conducted XRD measurements on the perovskite films. All authors analyzed their data. A.H. participated in the supervision of the work; M.S. directed and supervised the research; M.S. wrote the first draft of the paper; and all authors contributed to the discussion, commented, reviewed, and approved of the paper. **Competing interests:** None declared. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6413/449/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S26
Table S1
Reference (*40*)

20 February 2018; accepted 27 August 2018
10.1126/science.aat3583

MICROBIOTA

Transmission modes of the mammalian gut microbiota

Andrew H. Moeller[1,2,3]*†, Taichi A. Suzuki[2,3], Megan Phifer-Rixey[4,2,3], Michael W. Nachman[2,3]*

Mammals house a diversity of bacteria that affect health in various ways, but the routes by which bacterial lineages are transmitted between hosts remain poorly understood. We experimentally determined microbiota transmission modes by deriving 17 inbred mouse lines from two wild populations and monitoring their gut microbiotas for up to 11 host generations. Individual- and population-level microbiota compositions were maintained within mouse lines throughout the experiment, indicating predominantly vertical inheritance of the microbiota. However, certain bacterial taxa tended to be exchanged horizontally between mouse lines. Consistent with evolutionary theory, the degree of horizontal transmission predicted bacterial genera with pathogenic representatives responsible for human infections and hospitalizations.

All mammals associate with bacteria (*1*) in relationships that affect the hosts' digestive (*2*), immune (*3*), and neuroendocrine (*4*) systems. However, the routes by which specific bacterial lineages within the microbiota are transmitted between hosts remain poorly understood. Mammalian microbiotas are acquired both vertically from mother to offspring (*5*–*7*) and horizontally among non-kin through social interactions and shared environments (*8*–*10*). Quantifying the relative contributions of these transmission modes has been hindered by a lack of experiments designed to disentangle the transmission of individual bacterial lineages. Differentiating between vertical and horizontal transmission in humans, for example, is complicated by the difficulty of monitoring microbiotas for multiple host generations (*11*, *12*).

To differentiate between vertically and horizontally transmitted bacterial lineages in the mouse gut microbiota, we derived 17 mouse lines from wild populations in Tucson, AZ, and Edmonton, Alberta, Canada, and monitored their gut microbiotas during 3 years of inbreeding in a common laboratory environment (5 to 11 generations per line) (Fig. 1A and fig. S1). Mice from Tucson and from Edmonton were housed on different racks in the same room, and mice from different lines never came into direct contact. Every 1 to 2 weeks, mice were provided with new cages, food, and bedding. All cages were autoclaved, and food and bedding were sterilized. Cages were changed in random order with respect to population of origin. This experimental setup allowed us to quantify the fidelity of gut bacterial lineages to mouse lines and to identify gut bacterial lineages that were horizontally transmitted (through the mouse facility and animal handlers).

Cecal contents were extracted from wild-caught mice and subsequently from breeding individuals of each inbred line every generation, yielding a total of 212 samples (table S1). Illumina sequencing of the V4-V5 region of the 16S ribosomal DNA gene produced 15,353,040 sequences, averaging 72,420 reads per sample. To evaluate the relative contributions of vertical and horizontal transmission to the composition of the mouse microbiota, the binary Sorensen-Dice coefficient was calculated for every pairwise comparison between cecal samples across all mice. The binary Sorensen-Dice coefficient indicates the proportion of lineages that are shared by two communities and does not consider the relative abundances of lineages. Here, this coefficient is the proportion of amplicon sequence variants (ASVs) shared by two microbiotas. We defined binary Sorensen-Dice dissimilarity (BSDD) as one minus the binary Sorensen-Dice coefficient, such that a high BSDD indicates little overlap between the microbiotas' community memberships and a low BSDD indicates substantial overlap between the microbiotas' community memberships. Comparing BSDD within and between mouse lines allowed us to infer the contribution of vertical and horizontal transmission to the community membership of the mouse gut microbiota.

Comparing BSDDs within and between mouse lines indicated that vertical inheritance was the primary mode of gut bacterial transmission during the transition from the wild to the laboratory and subsequent inbreeding. Population- and individual-level compositional differences among wild-derived gut microbiotas were maintained within inbred mouse lines for 10 generations (Fig. 1). We found that Edmonton and Tucson mice that were sampled in the wild harbored compositionally distinct gut microbiotas [analysis of similarity (ANOSIM) P value = 0.001] (Fig. 1B). Edmonton and Tucson mice that were sampled in the laboratory maintained compositionally distinct gut microbiotas that reflected their wild population of origin (ANOSIM P value = 0.001) (Fig. 1C). The microbiotas of laboratory-born mice tended to be more similar to those of individuals sampled in their wild population of origin than they were to those of individuals sampled in the other wild population (t test P value = 3.15×10^{-23} for Edmonton, and t test P value = 7.58×10^{-7} for Tucson) (Fig. 1D). The distinctiveness of Edmonton and Tucson microbiotas was still evident in the 10th generation of the experiment (Fig. 1E). We observed that each mouse line, including sublines derived from the same wild-caught founders, maintained a compositionally distinct gut microbiota community membership throughout the experiment (Fig. 1F and table S2). The wild microbiota is not present in laboratory lines of mice that have been rederived under sterile conditions, such as the mouse model C57BL/6, but previous work has shown that germ-free mice inoculated with a wild gut microbiota can maintain the wild microbiota for four generations in gnotobiotic isolators (*12*). Our results indicate that individual-specific gut microbiota compositions of wild mice can be maintained in descendants for more than 10 generations in a mouse facility without gnotobiotic isolators.

Patterns of BSDD between samples indicated that vertical inheritance was the dominant mode of gut bacterial transmission, but they also indicated the horizontal transmission of some gut bacteria. Despite remaining distinct for 10 generations, the community memberships of the microbiotas of Tucson and Edmonton mice converged over the course of the experiment (Fig. 2A). We divided the experiment into seven time periods of equal length and found that BSDD between Tucson and Edmonton mice from the same period decreased with increasing time in the laboratory (nonzero linear coefficient, $P = 1.24 \times 10^{-8}$). This trend may indicate the exchange of bacteria between Tucson and Edmonton mice or parallel selection in the laboratory environment. To differentiate between these alternatives, we compared the microbiotas of later generations of laboratory-bred mice with the microbiotas of the founders of the lines derived from the other location (Edmonton for Tucson lines and Tucson for Edmonton lines). We found that their dissimilarity decreased with time (Fig. 2B). These results indicate that the gut microbiotas of laboratory-bred mice diverged from those of wild-caught mice initially, most likely due to a combination of drift and selection in the laboratory environment. Subsequently, the convergence of the gut microbiotas of laboratory-bred mice with those of wild-caught line founders from the other geographical origin indicated that gut bacteria were transmitted horizontally between lines derived from the different wild populations.

Given that the gut bacteria of wild-derived mouse lines displayed evidence of both vertical and horizontal transmission, we next investigated how specific bacterial genera were transmitted

[1]Miller Institute for Basic Research in Science, University of California, Berkeley, CA 94720, USA. [2]Museum of Vertebrate Zoology, University of California, Berkeley, CA 94720, USA. [3]Department of Integrative Biology, University of California, Berkeley, CA 94720, USA. [4]Department of Biology, Monmouth University, West Long Branch, NJ 07764, USA.
***Corresponding author. Email: andrew.moeller@berkeley.edu (A.H.M.); mnachman@berkeley.edu (M.W.N.)**
†Present address: Department of Ecology and Evolutionary Biology, Cornell University, New York, NY 14853, USA.

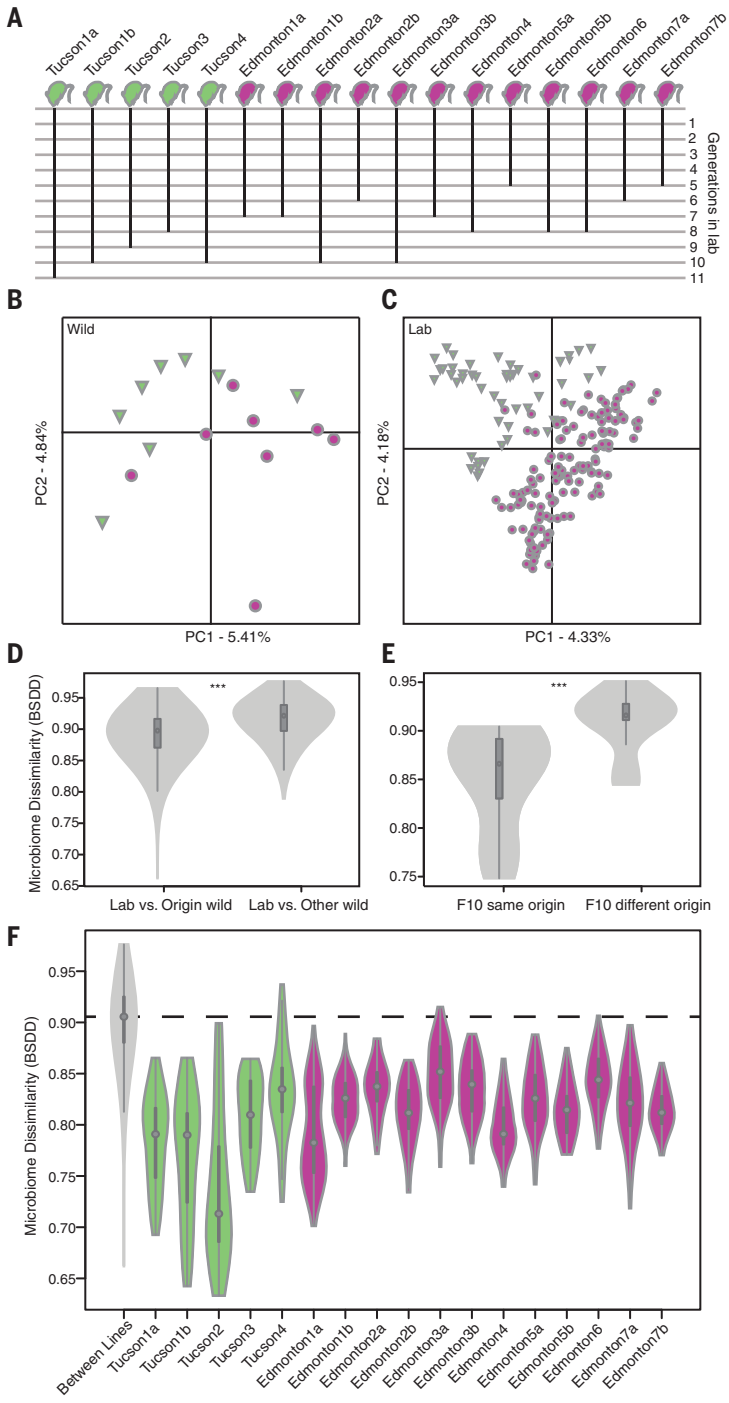


Fig. 1. Population- and individual-specific microbiotas of wild mice are vertically inherited for 10 generations in the laboratory. (**A**) Mouse cartoons and vertical lines represent inbred mouse lines. Sublines are labeled "a" and "b." Green and purple indicate Tucson and Edmonton, respectively. (**B**) Principal coordinates (PC) plot of BSDDs among mice sampled in the wild. (**C**) Principal coordinates plot of BSDDs among mice sampled in the laboratory. In (B) and (C), triangles and circles represent Tucson and Edmonton mice, respectively. (**D**) The violin plot on the left displays BSDDs between laboratory mice and wild mice from the same geographical origin. The violin plot on the right displays BSDDs between laboratory mice and wild mice from different geographical origins. (**E**) BSDDs among F_{10} mice descended from the same geographical origin (left) and from different geographical origins (right). Asterisks in (D) and (E) indicate significance; ***$P < 0.001$. (**F**) BSDDs among mice from different lines derived from the same geographical origin (gray), from the same Tucson line (green), and from the same Edmonton line (purple). The dashed line indicates the mean between-line BSDD. Rectangles in (D) to (F) delineate inner-quartile ranges. Circles in (F) indicate means.

between hosts. For each bacterial genus, we calculated the mean BSDD between mice from different lines and between mice from the same line. The ratio of between-line to within-line BSDD for each genus, which we call the transmission mode (TM) score, provided information about the degree of vertical versus horizontal transmission of ASVs belonging to the genus (fig. S2). A TM score of >1 indicated that ASVs of the genus tended to be restricted to specific mouse lines (i.e., were vertically inherited), a score equal to 1 indicated that ASVs were distributed equally among mice irrespective of line, and a score of <1 indicated that ASVs were more often shared by mice from different lines than by mice from the same line (i.e., were horizontally transmitted). TM scores of <1 were consistent with the rapid spread and decline of ASVs in the mouse colony (leading to increased similarity of the gut microbiotas of mice from different lines), as well as the epigenetic inheritance of mouse immune response [e.g., (*13*)] to specific ASVs (leading to decreased similarity of the gut microbiotas of mice from the same line). For example, bacterial genera within the Mollicutes (e.g., *Mycoplasma*) displayed TM scores of <1, and mice from different lines sampled in the same time period tended to share ASVs of these genera (fig. S3).

Comparing TM scores across bacterial genera produced a ranking of the fidelity of gut bacterial genera to mouse lineages (table S3). The majority of the bacterial genera displayed a tendency for vertical transmission. However, the distribution of TM scores across bacterial genera was bimodal (Hartigan's dip test P value = 0.001252), with a dip around 1 (Fig. 3A). Thus, bacterial genera in the mouse gut microbiota tend to be either exchanged horizontally or inherited vertically, with relatively few genera displaying mixed transmission strategies.

Transmission mode was associated with bacterial phylogenetic history. Three bacterial classes displayed a mean genus-level TM score significantly different from 1 (based on 95% confidence intervals): the Bacilli (0.954), the Clostridia (1.106), and the Bacteroidia (1.113) (Fig. 3B). In the Bacilli and the Clostridia, genera from the same class displayed TM scores that were more similar than were the TM scores of bacterial genera from different classes (F tests; $P = 0.0504$ for Bacilli, and $P = 0.00231$ for Clostridia). These results indicate that the two dominant classes of Firmicutes within the mammalian gut have evolved divergent transmission strategies. Horizontal transmission of the Bacilli and vertical inheritance of the Bacteroidia are evident in principal coordinates plots of the distributions of Bacilli and Bacteroidia ASVs within and between mouse lines (fig. S4).

Transmission routes were also associated with bacterial lifestyle (Fig. 3C and supplementary materials). Obligate anaerobes tended to be transmitted vertically (TM score > 1; t test P value = 0.01503), whereas obligate aerobes were transmitted horizontally (TM score < 1; t test P value = 0.01537) (Fig. 3C). These results suggest that oxygen tolerance enables the horizontal transmission

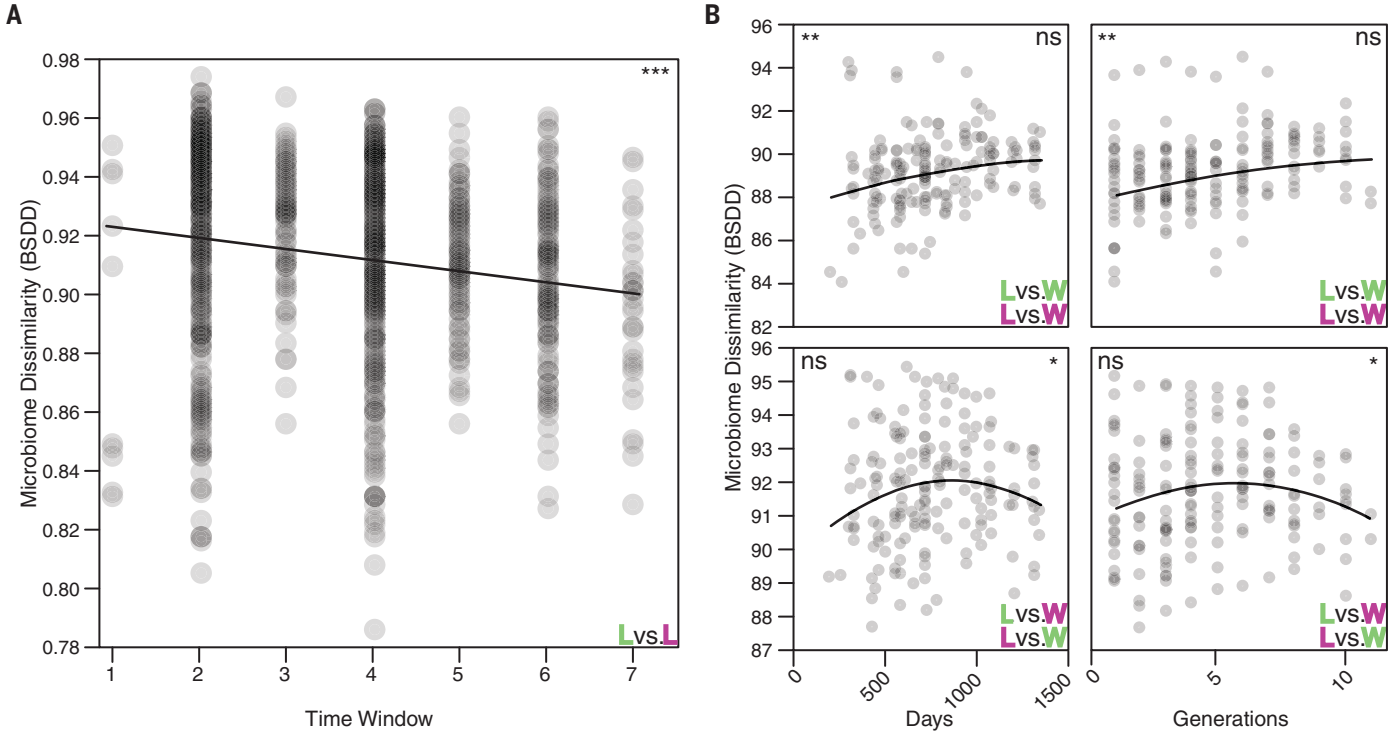


Fig. 2. Horizontal transmission of gut microbiota among mouse lines. (**A**) Dot plot displaying BSDDs between Tucson and Edmonton mice sampled in the same time period of the experiment. Each point represents a comparison between two laboratory mice descended from different geographical origins, denoted in the bottom left corner by the label "L vs. L," colored to correspond to Fig. 1. (**B**) Dot plots display the relationship between the number of days (left column of plots) or generations (right column of plots) in the laboratory and microbiota BSDDs between laboratory-bred mice and wild-caught line founders. The top row of plots displays comparisons between laboratory-bred mice and ancestral wild-caught line founders from the same geographical origin. The bottom row of plots displays comparisons between laboratory-bred mice and wild-caught line founders from the other geographical origin. "L" indicates laboratory-bred mice, "W" indicates wild-caught line founders, and colors correspond to Fig. 1. In (B), the curves indicate best-fit polynomial regressions, and the significance of nonzero linear and quadratic parameters is shown in the upper left and upper right corners of each plot, respectively. ns, not significant; *P < 0.05; **P < 0.01; ***P < 0.001.

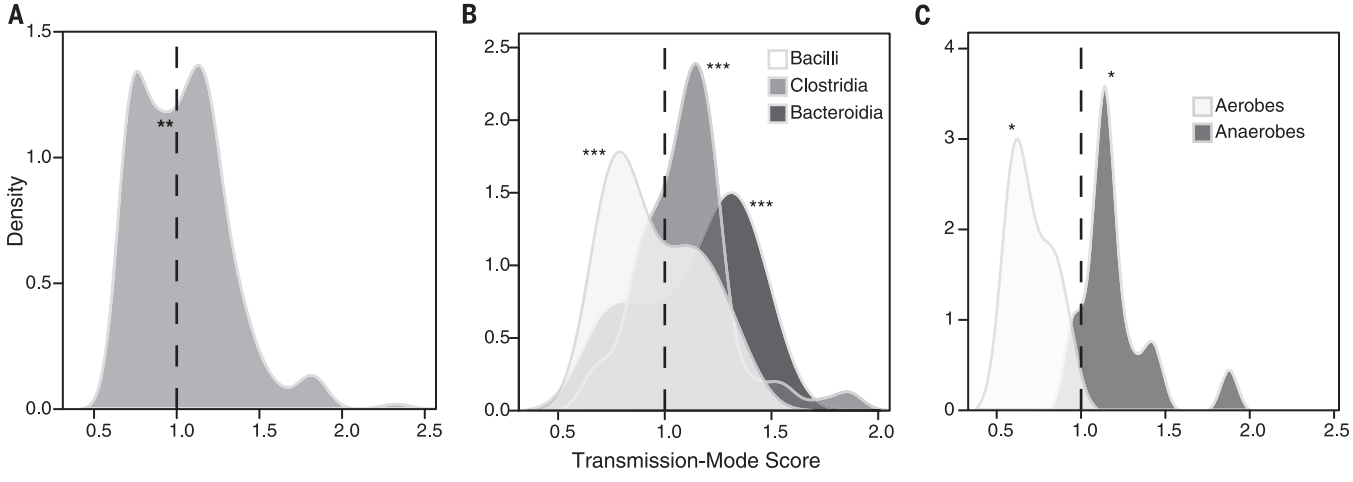


Fig. 3. Gut bacterial taxa follow divergent transmission strategies associated with evolutionary history and lifestyle. Kernal density plots show distributions of TM scores for (**A**) all bacteria; (**B**) the Bacilli, the Clostridia, and the Bacteroidia; and (**C**) obligate aerobes and obligate anaerobes. In (A) to (C), dashed lines denote a TM score of 1, and lighter to darker shades of gray indicate tendencies for horizontal and vertical transmission, respectively. The asterisks in (A) indicate the significance of Hartigan's dip test for unimodality. The distributions in (B) were all consistent with unimodality (Hartigan's dip test for unimodality, P > 0.05). Asterisks in (B) and (C) indicate the significance of t tests for mean TM scores different from 1. *P < 0.05; **P < 0.01; ***P < 0.001.

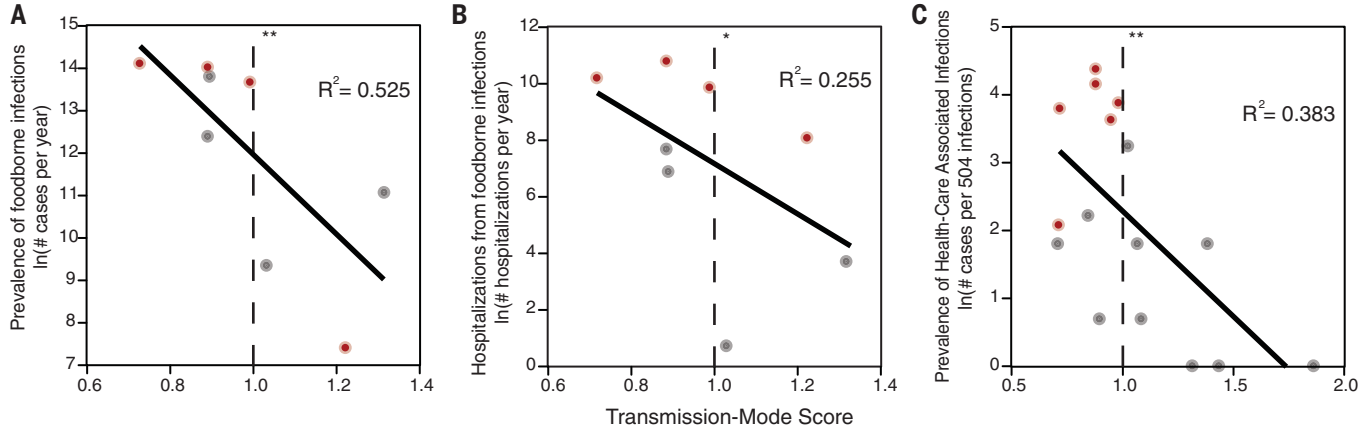


Fig. 4. TM scores predict infection and hospitalization rates for human pathogens. Dot plots display the relationships between the TM score and the number of foodborne infections (**A**), the number of hospitalizations from foodborne infections (**B**), and the number of HAIs (**C**) in the United States caused by bacterial genera as reported by (*18*) and (*19*). Red points indicate bacteria monitored by the Centers for Disease Control and Prevention's Emerging Infections Program. In (A) to (C), each point represents a bacterial genus, dashed lines denote a TM score of 1, and trend lines indicate linear regressions, with the significance of the nonzero slope denoted by * ($P < 0.05$) and ** ($P < 0.01$). R^2, coefficient of determination.

of gut bacterial lineages. However, spore-forming bacterial genera did not display any tendency for vertical or horizontal transmission ($P > 0.05$) (fig. S5).

Evolutionary theory suggests that transmission mode influences the evolution of virulence, with horizontal transmission favoring increased virulence relative to strict vertical transmission (*14*–*18*). To examine the relationship between TM score and virulence, we obtained epidemiological data on the number and severity of hospital-associated and foodborne infections in humans caused by gut bacterial genera in the United States. The genera currently monitored by the Centers for Disease Control and Prevention's Emerging Infections Programs displayed a mean TM score significantly less than 1 (Fig. 4) (*t* test *P* value = 0.0345). Moreover, the TM score was significantly associated with the number of infections and hospitalizations in humans caused by different genera. Linear regression analyses indicated that the TM score explained 52, 25, and 38% of the variance among genera in the rate at which they cause foodborne infections, hospitalizations from foodborne infections, and health care–associated infections (HAIs), respectively, in humans in the United States, on the basis of epidemiological data [(table 2 of (*19*) and table 3 of (*20*)]. Therefore, bacterial pathogens in humans belong to genera that appear to be adapted for transmission through the indoor environment. Previous work has shown that pathogenic bacteria are often generalists capable of persisting in both host-associated and external environments (*21*). Our results suggest that horizontally transmitted bacterial genera in mice (table S3) and humans are more likely to exhibit virulence than are vertically transmitted bacterial genera.

We conducted a long-term, multigenerational assessment of the transmission modes of bacterial genera in the mammalian gut microbiota. Our data showed that the majority of the murine gut microbiota was vertically inherited, such that the compositional identities of the gut microbiotas of individual mice were maintained for 10 generations of inbreeding in laboratory conditions. We also discovered that a minority of the gut microbiota was horizontally transmitted, leading to convergence in gut microbiota composition among mouse lines over time. Consistent with evolutionary theory, we found that the propensity for horizontal transmission of bacterial genera in mice is associated with the pathogenicity of the genera in humans, on the basis of epidemiological data (*19*, *20*).

REFERENCES AND NOTES

1. M. McFall-Ngai *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 3229–3236 (2013).
2. V. Tremaroli, F. Bäckhed, *Nature* **489**, 242–249 (2012).
3. J. L. Round, S. K. Mazmanian, *Nat. Rev. Immunol.* **9**, 313–323 (2009).
4. R. Diaz Heijtz *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3047–3052 (2011).
5. P. A. Vaishampayan *et al.*, *Genome Biol. Evol.* **2**, 53–66 (2010).
6. M. C. Collado, S. Delgado, A. Maldonado, J. M. Rodríguez, *Lett. Appl. Microbiol.* **48**, 523–528 (2009).
7. M. G. Dominguez-Bello *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 11971–11975 (2010).
8. J. Tung *et al.*, *eLife* **4**, e05224 (2015).
9. A. H. Moeller *et al.*, *Sci. Adv.* **2**, e1500997 (2016).
10. A. C. Perofsky, R. J. Lewis, L. A. Abondano, A. Di Fiore, L. A. Meyers, *Proc. R. Soc. London Ser. B* **284**, 20172274 (2017).
11. E. D. Sonnenburg *et al.*, *Nature* **529**, 212–215 (2016).
12. S. P. Rosshart *et al.*, *Cell* **171**, 1015–1028.e13 (2017).
13. R. M. Gorczynski, E. J. Steele, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2871–2875 (1980).
14. M. J. Wade, *Nat. Rev. Genet.* **8**, 185–195 (2007).
15. M. Bright, S. Bulgheresi, *Nat. Rev. Microbiol.* **8**, 218–230 (2010).
16. I. Pagán, N. Montes, M. G. Milgroom, F. García-Arenal, *PLOS Pathog.* **10**, e1004293 (2014).
17. J. C. de Roode, A. J. Yates, S. Altizer, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 7489–7494 (2008).
18. M. Lipsitch, S. Siller, M. A. Nowak, *Evolution* **50**, 1729–1741 (1996).
19. E. Scallan *et al.*, *Emerg. Infect. Dis.* **17**, 7–15 (2011).
20. S. S. Magill *et al.*, *N. Engl. J. Med.* **370**, 1198–1208 (2014).
21. S. P. Brown, D. M. Cornforth, N. Mideo, *Trends Microbiol.* **20**, 336–342 (2012).

ACKNOWLEDGMENTS

We thank G. Heyer for aiding in animal husbandry and B. Koskella and C. Brook for comments on the manuscript. **Funding:** This work was supported by NIH RO1 GM074245 to M.W.N. and by a Miller research fellowship to A.H.M. **Author contributions:** A.H.M. designed the study, conducted analyses, and wrote the manuscript; T.A.S. and M.P.-R. performed experiments and edited the manuscript; M.W.N. designed the study, oversaw the creation of mouse inbred lines, and wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data included in this study are available in the European Nucleotide Archive under accession number PRJEB28216.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6413/453/suppl/DC1
Materials and Methods
Figs. S1 to S5
Tables S1 to S3
References (*22*–*25*)

27 March 2018; resubmitted 18 June 2018
Accepted 10 September 2018
10.1126/science.aat7164

EVOLUTION

Agouti-related peptide 2 facilitates convergent evolution of stripe patterns across cichlid fish radiations

Claudius F. Kratochwil[1,2,3]*, Yipeng Liang[1], Jan Gerwin[1,3], Joost M. Woltering[1], Sabine Urban[1,3], Frederico Henning[1,4], Gonzalo Machado-Schiaffino[1,5], C. Darrin Hulsey[1,3], Axel Meyer[1,3]*

The color patterns of African cichlid fishes provide notable examples of phenotypic convergence. Across the more than 1200 East African rift lake species, melanic horizontal stripes have evolved numerous times. We discovered that regulatory changes of the gene *agouti-related peptide 2* (*agrp2*) act as molecular switches controlling this evolutionarily labile phenotype. Reduced *agrp2* expression is convergently associated with the presence of stripe patterns across species flocks. However, cis-regulatory mutations are not predictive of stripes across radiations, suggesting independent regulatory mechanisms. Genetic mapping confirms the link between the *agrp2* locus and stripe patterns. The crucial role of *agrp2* is further supported by a CRISPR-Cas9 knockout that reconstitutes stripes in a nonstriped cichlid. Thus, we unveil how a single gene affects the convergent evolution of a complex color pattern.

Stephen Jay Gould famously posited that if it were possible to rerun the "tape of life," outcomes would be different (*1*). The relative importance of determinism and contingency during evolution is still far from settled (*2*, *3*). But for particular groups of organisms, one can now test Gould's hypothesis. For instance, in less than 8 million to 12 million years, more than 1200 species of cichlid fishes have evolved to form repeated adaptive radiations in the East African Rift Valley lakes, such as Lakes Victoria, Tanganyika, and Malawi (Fig. 1B) (*4*–*8*). These adaptive radiations have given rise to a large diversity of species displaying various color patterns (Fig. 1, C to N), including the repeated occurrence of melanic horizontal stripes (Fig. 1A and supplementary text). Convergent evolution is prevalent in the East African cichlid radiations (*9*–*11*), providing a replicated natural experiment whereby distantly related species from independent adaptive radiations can be used to determine what mechanisms have generated these recurrent phenotypes (*12*–*16*). More specifically, we address whether horizontal stripes, a convergent phenotype, have an identical, similar, or different molecular bases across the independent adaptive radiations of cichlid fishes.

Previously (*17*)—using a genetic mapping panel of two Lake Victoria species, *Pundamilia nyererei* (*Pnye*, nonstriped, Fig. 1E) and *Haplochromis sauvagei* (*Hsau*, striped, Fig. 1C)—we found that horizontal stripes (Fig. 2C) are inherited as a recessive Mendelian trait mapping to chromosome 18 (Fig. 2A). This was confirmed by a second cross involving the same nonstriped species and another striped species, *H. chilotes* (*Hchi*, striped) (Fig. 2A and supplementary text). To more precisely isolate the causal genetic interval for stripe presence, we fine-mapped the trait using recombinant F_2 individuals of the *Pnye* × *Hsau* cross and reduced the causal interval from 600 to 25 kb (fig. S1). This interval contained the genes *agouti-related peptide 2* (*agrp2*), *v-type proton ATPase subunit d 2* (*atp6V0d2*), and an unknown gene (*unk*) (Fig. 2B). The resequencing of all coding regions revealed no fixed missense or nonsense mutations (fig. S2), suggesting that cis-regulatory variation determines stripe presence.

The teleost-specific *agrp2* (fig. S3) is a strong candidate gene for stripes because its paralogs have been previously associated with pigmentation phenotypes (*18*–*20*). To test for *agrp2* expression differences between nonstriped (*Pnye*) and striped (*Hsau*) Lake Victoria cichlids, we performed quantitative polymerase chain reaction (qPCR; Fig. 2D and fig. S5) on a number of adult tissues, including skin (supplementary text). Here, *agrp2* showed a significantly higher expression in the skin of *Pnye* (Fig. 2D and fig. S4). The lack of consistent expression variation between melanic and nonmelanic regions and generally across dorsoventral and anterior-posterior positions suggests that *agrp2* does not shape pigmentation patterns through local expression-level variation but rather acts as a general stripe pattern inhibitor (fig. S6). Whereas qPCR revealed no such expression differences for paralogs and neighboring genes (supplementary text), qPCRs on F_2 *Pnye*/*Hsau* hybrid individuals confirmed that expression differences are linked to the *agrp2* locus and exhibit an allelic dosage effect as expected for cis-regulatory mutations (fig. S4).

To identify causal mutations affecting both *agrp2* expression and stripe phenotype, we sequenced the *agrp2* locus in individuals ($n \geq 10$ individuals per population) from natural populations of the three hybrid-cross species. We screened for alternatively fixed, fully associated variants with the stripe phenotype in pairwise comparisons of each striped species (*Hsau* or *Hchi*) versus the nonstriped *Pnye*. Our analyses indicated that a 1.1-kb interval within the first *agrp2* intron (Fig. 2B) that exhibited shared, alternatively fixed alleles is a strong candidate region for a regulatory element controlling *agrp2* expression (fig. S7). To test whether this 1.1-kb interval [*enhancer of agrp2 in Pnye* (*Pnye.enh.a*)] contains cis-regulatory elements that could influence interspecific differences between striped and nonstriped species, we tested the elements of both species in a green fluorescent protein (GFP) reporter assay in vivo (supplementary text). It showed that *Pnye.enh.a* efficiently modulates GFP expression [Tukey's honest significant difference (HSD), $P < 0.001$] and is significantly more potent than the homologous sequence of the striped species *Hsau* (Tukey's HSD, $P < 0.001$) (Fig. 2, E and F, and fig. S8). Together, these results indicate that higher expression of *agrp2*, and thereby the suppression of stripe patterns, is indeed enhanced by *Pnye.enh.a* (fig. S9).

Our results reveal *agrp2* as a major determinant of stripe presence that might be sufficient to suppress stripe patterns in *Pnye*. To further test this finding, we used CRISPR-Cas9 genome engineering to manipulate *agrp2* and to thereby potentially derepress stripe patterns. *Pnye* eggs were injected with Cas9 and *agrp2* guide RNAs, and we obtained four mutants, all of which had nonsense and frameshift mutations within *agrp2* (fig. S10 and table S1). These CRISPR-Cas9 mutants developed a continuous midlateral stripe (Fig. 2H and fig. S10) yet no dorsolateral stripe (supplementary text). Because horizontal stripes were never observed in noninjected *Pnye* individuals (>100 observations; Fig. 2G), this strongly suggests that although species such as *Pnye* have no stripes, the genomic and developmental machinery for stripe pattern formation is in place, and stripes can reappear in this nonstriped species by experimental manipulation of *agrp2*.

Next, we tested if the expression levels of *agrp2* and stripe patterns are generally associated across other cichlid species from the repeated species flocks of Lakes Victoria, Malawi, and Tanganyika, suggesting a shared molecular basis for convergent stripe phenotypes. Using qPCR on adult skins of striped and nonstriped species of each of the three major East African cichlid radiations (in total, 24 species; fig. S11), we revealed that nonstriped species commonly had higher *agrp2* expression levels than striped species (Fig. 3B). This association was confirmed

[1]Department of Biology, University of Konstanz, Konstanz, Germany. [2]Zukunftskolleg, University of Konstanz, Konstanz, Germany. [3]International Max Planck Research School for Organismal Biology (IMPRS-OB), Max Planck Institute for Ornithology, Konstanz, Germany. [4]Department of Genetics, Institute of Biology, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil. [5]Department of Functional Biology, Area of Genetics, University of Oviedo, Oviedo, Spain.
*Corresponding author. Email: claudius.kratochwil@uni-konstanz.de (C.F.K.); axel.meyer@uni-konstanz.de (A.M.)

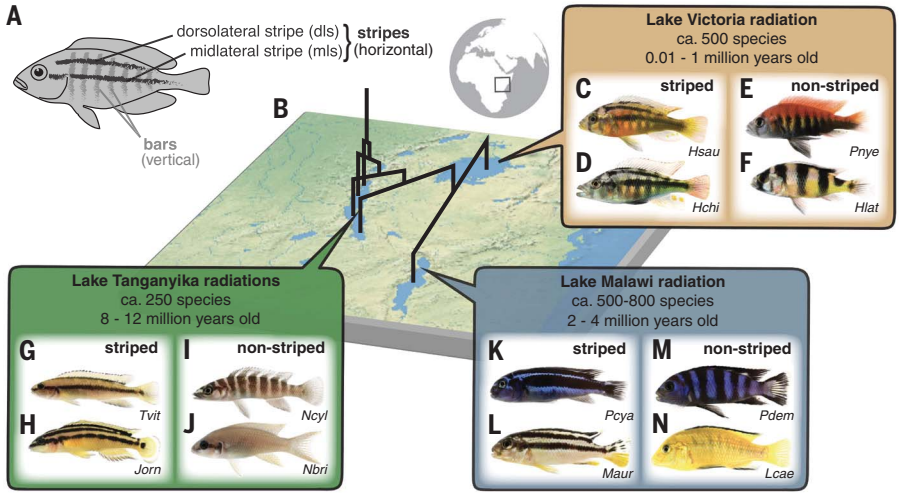


Fig. 1. Convergent evolution of horizontal stripes across African cichlid radiations. (**A**) Schematic of melanic horizontal stripes and vertical bars. (**B**) Map of the African Great Lakes Victoria, Tanganyika, and Malawi and superimposed simplified phylogenetic tree of the adaptive radiations of Lakes Tanganyika (green; not all paraphyletic lineages shown), Malawi (blue), and Victoria (orange). (**C** to **N**) Twelve of the focal striped and nonstriped species of this study, including species from Lakes Victoria [(C) to (F)], Tanganyika [(G) to (J)], and Malawi [(K) to (N)]. *Hlat*, *Haplochromis latifasciatus*; *Jorn*, *Julidochromis ornatus*; *Lcae*, *Labidochromis caeruleus*; *Maur*, *Melanochromis auratus*; *Nbri*, *Neolamprologus brichardi*; *Ncyl*, *Neolamprologus cylindricus*; *Tvit*, *Telmatochromis vittatus*.

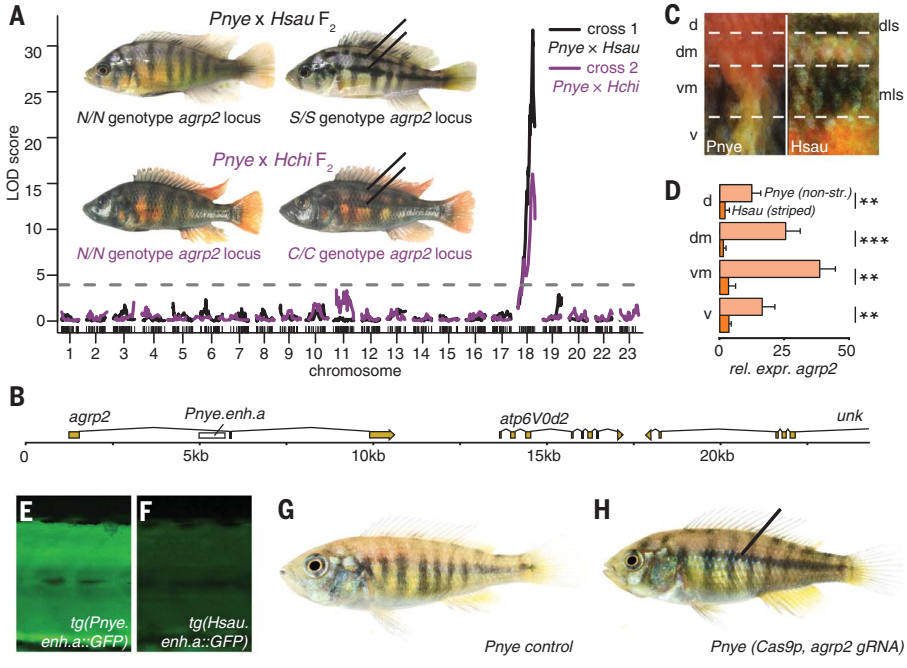


Fig. 2. *Agrp2* controls stripe loss in Lake Victoria cichlids. (**A**) In two mapping crosses between *Pnye* and *Hsau* and *Pnye* and *Hchi*, horizontal stripes map to the same region on chromosome 18. (**B**) F_2 recombinant fine-mapping isolates a 25-kb interval containing *agrp2*. (**C** and **D**) Skin biopsies show differential *agrp2* expressions between skins from *Hsau* (C) and *Pnye* (D). Error bars indicate means + SD. (**E** and **F**) An intronic 1.1-kb element of *Pnye* (*Pnye.enh.a*) is regulatory active (E) and shows stronger activity than the homologous sequence of the striped species *Hsau* (F) in a zebrafish larvae GFP reporter assay. (**G** and **H**) CRISPR-Cas9 mediated knockouts of *agrp2* in the normally nonstriped cichlid *Pnye* (G) develop stripes [(H); midlateral stripe indicated by line]. LOD score, logarithm of the odds score; *N/N*, homozygous for the *Pnye agrp2* allele; *S/S*, homozygous for the *Hsau agrp2* allele; *C/C*, homozygous for the *Hchi agrp2* allele; *tg*, transgenic construct; gRNA, guide RNA; mls, midlateral stripe; dls, dorsolateral stripe; d, dorsal; dm, dorsomedial; vm, ventromedial; v, ventral.

by comparative phylogenetic analyses (Fig. 3A), demonstrating a significant evolutionary association between low *agrp2* expression and stripe presence [phylogenetic analysis of variance (ANOVA); mean $P < 0.001$; supplementary text].

To determine if this convergence at the phenotypic and *agrp2* gene expression level is also paralleled at the sequence level (*16*), we comparatively analyzed homologous *enh.a* sequences across cichlids from Lakes Victoria, Malawi, and Tanganyika. A tree of *enh.a* revealed substantial sequence variation and resolved striped Lake Victoria species as monophyletic, suggesting a single origin of the striped alleles, whereas striped species of other lakes were not monophyletic (Fig. 3C). None of the nine mutations within *enh.a* that showed complete association with stripes in Lake Victoria cichlids showed similar stripe association in cichlids of Lakes Malawi or Tanganyika (figs. S12 and S13). Consequently, independent mutations must be affecting *agrp2* expression and thereby stripe patterns across the three major cichlid radiations (Fig. 4D).

Lastly, we tested whether the same locus is responsible for stripe-pattern variation outside of Lake Victoria cichlids using a hybrid cross between the nonstriped Lake Malawi species *Pseudotropheus demasoni* (*Pdem*; Fig. 1M) and striped species *Ps. cyaneorhabdos* (*Pcya*; Fig. 1K) that also differed in their skin *agrp2* expression (fig. S14). We obtained 270 F_2 hybrid individuals that were genotyped at the *agrp2* locus and phenotyped regarding their stripe patterns (Fig. 4, B and C; fig. S15; and supplementary text). The results revealed significant linkage between the *agrp2* allele and stripe presence (Fisher's exact test, $P = 7.6 \times 10^{-8}$; table S2). The allelic variation at the *agrp2* locus explains more than 50% of the phenotypic variance in stripe patterns [Cox-Snell or Nagelkerke pseudo-R^2 from ordered logistic regression; table S3]. Nevertheless, the phenotypic distribution of F_2 individuals (49 nonstriped and 221 striped individuals) differed from the Mendelian 3:1 ratio observed in the Lake Victoria crosses (chi-square test, $P < 0.001$), providing evidence for additional minor modifier loci. These results strongly suggest that *agrp2* acts as a major determinant of stripe pattern absence or presence in Lake Malawi (as in the younger Lake Victoria radiation), but additional minor stripe modifiers have evolved or were recruited in the older Lake Malawi radiation (Fig. 4D).

The repeated evolution of horizontal stripes in East African cichlid radiations is facilitated by cis-regulatory evolution of *agrp2*. Despite its described role and function in the brain (*18*), we have discovered a hitherto unknown function for this gene in the skin that highlights notable functional similarities between Agrp2 and the mammalian Agouti (Asip) as well as teleost Asip1 (*19*, *20*). From what is known about proteins of the Agouti family, Agrp2 likely acts as an antagonist for the melanocortin receptors Mc1r and/or Mc5r (*21*). Low Agrp2 levels would trigger stripe melanophore proliferation, pigment

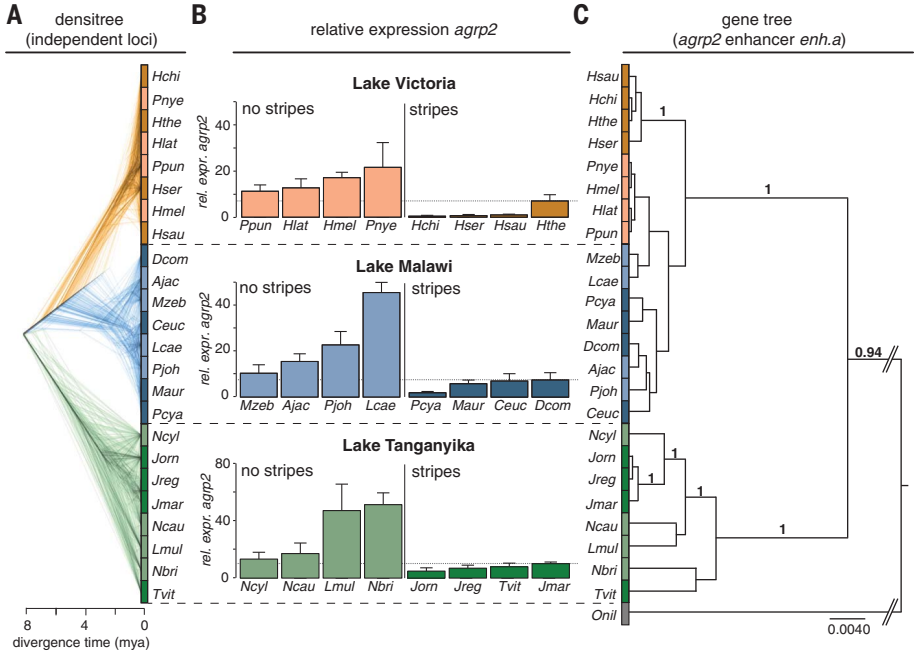


Fig. 3. Regulatory changes of *agrp2* are predictive of convergent stripe evolution. (**A**) Densitree representation of phylogeny for the 24 examined species, including divergence time estimates (*8*). mya, million years ago. (**B**) Skin gene expression analysis of *agrp2* across adaptive radiations highlights the strong evolutionary association of stripes with low *agrp2* expression levels (phylogenetic ANOVA, $P < 0.001$). Error bars indicate means + SD (**C**) A gene (locus) tree of the cis-regulatory element *enh.a* supports a single origin of striped alleles in Lake Victoria. Numbers present posterior probabilities >0.9. *Ajac, Aulonocara jacobfreibergi*; *Ceuc, Cheilochromis euchilus*; *Dcom, Dimidiochromis compressiceps*; *Hmel, Haplochromis melanopterus*; *Hser, Haplochromis serranus*; *Hthe, Haplochromis thereuterion*; *Jmar, Julidochromis marlieri*; *Jreg, Julidochromis regani*; *Lmul, Lamprologus multifasciatus*; *Mzeb, Maylandia zebra*; *Ncau, Neolamprologus caudopunctatus*; *Onil, Oreochromis niloticus*; *Pjoh, Placidochromis johnstoni*; *Ppun, Pundamilia pundamilia.*

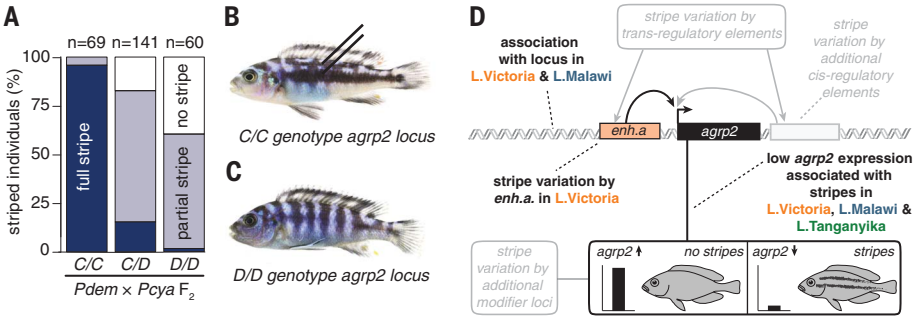


Fig. 4. A shared genomic basis of stripes across cichlid radiations. (**A**) Stripes are associated with *agrp2* alleles in Lake Malawi hybrid F_2 individuals (*Pdem* × *Pcya*). Inheritance is not Mendelian, suggesting that additional genetic modifiers exist in Lake Malawi cichlids. (**B** and **C**) F_2 hybrids homozygous for the *Pcya agrp2* allele (*C/C*, showing stripes, black lines) and *Pdem* allele (*D/D*, no stripes) exhibit clear stripe pattern differences. Parental species are shown in Fig. 1K (striped) and Fig. 1M (nonstriped). (**D**) Summary of the known (black) and unknown (gray) aspects of the genetic control of cichlid horizontal stripes.

dispersion, and/or pigment production (stripe patterns present), whereas high levels would block these processes (no stripe patterns) (*21*). Expression levels of *agrp2* thereby act as a switch controlling stripe presence and absence. In Lake Victoria, expression-level differences seem to be caused by several mutations in a 1.1-kb intronic regulatory region (*enh.a*; Fig. 2B) that push the expression of *agrp2* levels above or below a threshold that determines the stripe phenotype. Such a threshold-based molecular on-off switch may have permitted the frequent loss as well as reevolution of stripes within East African cichlids. Although the presence of stripes appears to be controlled by differential expression of the same gene (*agrp2*), causal genetic variants must differ among the independent radiations of Lakes Victoria, Malawi, and Tanganyika (Fig. 4D and figs. S12 and S13). The intermediate phenotypes obtained from the Malawi cross (Fig. 4A), together with the lack of the dorsolateral stripe in the CRISPR-Cas9 mutants (Fig. 2H), provide evidence for additional modifier loci determining stripe presence (Fig. 4D). However, those seem generally less prominent in the young (<15,000 years old) Lake Victoria radiation compared to the older (2 million to 4 million years old) Lake Malawi radiation (supplementary text).

Regulatory variation of *agrp2* provides a molecular basis for the repeated evolution and loss of stripe patterns across cichlid species flocks. Recurrent regulatory evolution at the *agrp2* locus constitutes an example of regulatory tinkering (*22*, *23*) that might have facilitated the ease and speed of the evolution of both converged and diverged phenotypes that characterizes the East African cichlid radiations. The simplicity of such a threshold mechanism might have permitted the phylogenetically observed rapid losses and reevolutions of stripe patterns. Therefore, Stephen Jay Gould's predictions (*1*) appear questionable at this evolutionary scale, and if one were to replay the evolution of cichlid adaptive radiations, the results might be surprisingly similar: striped and nonstriped cichlids evolving again and again through regulatory evolution at the *agrp2* locus.

REFERENCES AND NOTES

1. S. J. Gould, *Wonderful Life: The Burgess Shale and the Nature of History* (W. W. Norton & Company, 1990).
2. J. B. Losos, T. R. Jackman, A. Larson, K. Queiroz, L. Rodriguez-Schettino, *Science* **279**, 2115–2118 (1998).
3. J. B. Losos, *Improbable Destinies* (Penguin, 2017).
4. T. D. Kocher, *Nat. Rev. Genet.* **5**, 288–298 (2004).
5. A. Meyer, T. D. Kocher, P. Basasibwaki, A. C. Wilson, *Nature* **347**, 550–553 (1990).
6. G. F. Turner, *Curr. Biol.* **17**, R827–R831 (2007).
7. M. E. Santos, W. Salzburger, *Science* **338**, 619–621 (2012).
8. D. Brawand *et al.*, *Nature* **513**, 375–381 (2014).
9. O. Seehausen, P. J. Mayhew, J. J. M. Van Alphen, *J. Evol. Biol.* **12**, 514–534 (1999).
10. M. Muschick, A. Indermaur, W. Salzburger, *Curr. Biol.* **22**, 2362–2368 (2012).
11. A. Meyer, *Trends Ecol. Evol.* **8**, 279–284 (1993).
12. S. Kuraku, A. Meyer, *Curr. Opin. Genet. Dev.* **18**, 551–558 (2008).
13. M. E. Protas, N. H. Patel, *Annu. Rev. Cell Dev. Biol.* **24**, 425–446 (2008).
14. R. B. Roberts, J. R. Ser, T. D. Kocher, *Science* **326**, 998–1001 (2009).
15. C. F. Kratochwil, A. Meyer, *BioEssays* **37**, 213–226 (2015).
16. D. L. Stern, *Nat. Rev. Genet.* **14**, 751–764 (2013).
17. F. Henning, H. J. Lee, P. Franchini, A. Meyer, *Mol. Ecol.* **23**, 5224–5240 (2014).
18. C. Zhang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 20164–20171 (2010).
19. M. Manceau, V. S. Domingues, R. Mallarino, H. E. Hoekstra, *Science* **331**, 1062–1065 (2011).
20. R. M. Ceinos, R. Guillot, R. N. Kelsh, J. M. Cerdá-Reverter, J. Rotllant, *Pigment Cell Melanoma Res.* **28**, 196–209 (2015).
21. L. Cal, P. Suarez-Bregua, J. M. Cerdá-Reverter, I. Braasch, J. Rotllant, *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **211**, 26–33 (2017).
22. C. F. Kratochwil, A. Meyer, *Curr. Biol.* **25**, R285–R288 (2015).
23. A. Mazo-Vargas *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 10701–10706 (2017).

24. C. F. Kratochwil *et al*., Data from: Agouti-related peptide 2 facilitates convergent evolution of stripe patterns across cichlid fish radiations. Dryad (2018); doi:10.5061/dryad.h1107nd.

ACKNOWLEDGMENTS

The authors thank S. Finkernagel for her exceptional technical assistance; A. Kautt, A. Nater, J. Torres-Dowdall, and the anonymous referees for insightful comments that improved the manuscript; and the University of Konstanz bioimaging center and animal facility. **Funding:** This work was supported by the Swiss National Science Foundation (P2BSP3_148629), the Marie Curie Zukunftskolleg Incoming Fellowship Program (291784), the Baden-Württemberg Foundation, the Deutsche Forschungsgemeinschaft (DFG, KR 4670/2-1), and the Young Scholar Fund of the University of Konstanz (all to C.F.K.). Y.L. is funded by the China Scholarship Council (CSC). A.M. and C.D.H. are funded by DFG and University of Konstanz. Partial funding came from an ERC Advanced Grant GenAdap (no. 293700) to A.M. **Author contributions:** C.F.K. conceptualized the study; provided funding; planned and supervised all experiments; collected, analyzed, and visualized all data, if not otherwise indicated; and wrote the manuscript. Y.L. performed qPCRs and in situ hybridization. J.G. and S.U. conducted the Lake Malawi cross and assisted with the sequencing and enhancer assay. J.M.W. designed and generated the CRISPR-Cas9 knockouts and helped with the enhancer assay. F.H. and G.M.-S. provided data for Lake Victoria crosses and contributed to their analysis. C.D.H. conducted the phylogenetic analysis and edited the manuscript. A.M. conceptualized the study, provided funding, and revised the manuscript. All authors gave comments on the manuscript and approved the final version. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Additional data have been deposited on Dryad (*24*) and are available in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6413/457/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S15
Tables S1 to S5
References (*25*–*63*)

16 October 2017; resubmitted 19 April 2018
Accepted 12 September 2018
10.1126/science.aao6809

PALEONTOLOGY

The nearshore cradle of early vertebrate diversification

Lauren Sallan[1]*†, Matt Friedman[2], Robert S. Sansom[3], Charlotte M. Bird[4], Ivan J. Sansom[4]*†

Ancestral vertebrate habitats are subject to controversy and obscured by limited, often contradictory paleontological data. We assembled fossil vertebrate occurrence and habitat datasets spanning the middle Paleozoic (480 million to 360 million years ago) and found that early vertebrate clades, both jawed and jawless, originated in restricted, shallow intertidal-subtidal environments. Nearshore divergences gave rise to body plans with different dispersal abilities: Robust fishes shifted shoreward, whereas gracile groups moved seaward. Fresh waters were invaded repeatedly, but movement to deeper waters was contingent upon form and short-lived until the later Devonian. Our results contrast with the onshore-offshore trends, reef-centered diversification, and mid-shelf clustering observed for benthic invertebrates. Nearshore origins for vertebrates may be linked to the demands of their mobility and may have influenced the structure of their early fossil record and diversification.

The ancestral habitat of vertebrates has long been debated, with opinions ranging from freshwater to open ocean habitats (*1*–*3*). Inferences have been derived from either the evolutionarily distant modern fauna or qualitative narratives based on select fossils. Early records of vertebrate divisions, such as jawed fishes and their relatives (total-group gnathostomes), consist of long gaps between inferred origination and definitive appearances (ghost lineages), punctuated by suggestive microfossils (*4*–*7*). Vertebrates, apart from toothlike conodont elements, were restricted in Ordovician ecosystems as trivial components of the Great Ordovician Biodiversification Event (*4*, *5*, *7*). Ancestral habitat is a critical factor in determining both pattern and mode of diversification, potential mismatches between biodiversity and available habitat area, and the source of apparent relationships with changing sea level (*6*). A lack of early vertebrate fossil data and habitat information in compendia has limited quantitative approaches (*4*), preventing resolution of this outstanding issue in vertebrate evolution.

We developed a database of total-group gnathostome occurrences (~480 million to 360 million years ago) (*4*, *5*, *8*) during their mid-Paleozoic diversification (n = 2827) (*9*) (fig. S1). Data collection focused on all occurrences from the interval encompassing the five oldest localities for each major clade (n = 188) (Fig. 1 and figs. S1 and S2) and phylogenetically constrained genera within jawless groups (n = 785) (Figs. 2 and 3 and figs. S1, S3, and S4) for use with Bayesian ancestral state reconstruction. We used environmental, lithological, and invertebrate community information from the literature and available databases to assign occurrences to benthic assemblage zones (*10*) (Fig. 1). Benthic assemblage zones are categorized and ordered as fresh water (BA0); intertidal above typical wave base (BA1); shallow subtidal and/or lagoon (BA2); deeper subtidal, including the start of tabulate coral–stromatoporoid reef systems (BA3); middle to outer shelf (BA4 and BA5); and shelf margin toward the bathyal region (BA6). These zones have been widely used in studies of mid-Paleozoic paleocommunities (*1*, *10*–*12*) (Fig. 1).

We applied Bayesian threshold models to phylogenies of occurrences using prior probabilities of residence in each benthic assemblage zone. This methodology allowed positive inference of both ancestral habitats and amount of evolutionary change required to move between zones ("liability" values) (*13*). All major clades, from the first skeletonizing jawless fishes (astraspids, arandaspids) to jawed bony fishes (osteichthyans), originated within nearshore intertidal and subtidal zones (~BA1 to BA3), centered on BA2, over a period of more than 100 million years (Fig. 1A and fig. S3). This area is relatively shallow, includes lagoons in reefal systems, and is located entirely above the storm wave base in the mid-Paleozoic (*11*) (Fig. 1).

We appraised whether nearshore origination in gnathostomes resulted from environmental bias in the record through comparison with habitat distributions for other facets of the mid-Paleozoic captured in independent datasets, including fossiliferous strata, regional paleocommunities, and global occurrences and richness (number of genera) (Fig. 1B and figs. S11 to S16) (*10*, *14*). Analysis of mid-Paleozoic strata in the Paleobiology Database (PBDB) (*14*), binned by distinct habitat categories (n = 4437), produced a distribution clustered on deep subtidal and reef environments [equivalent to BA3 and BA4

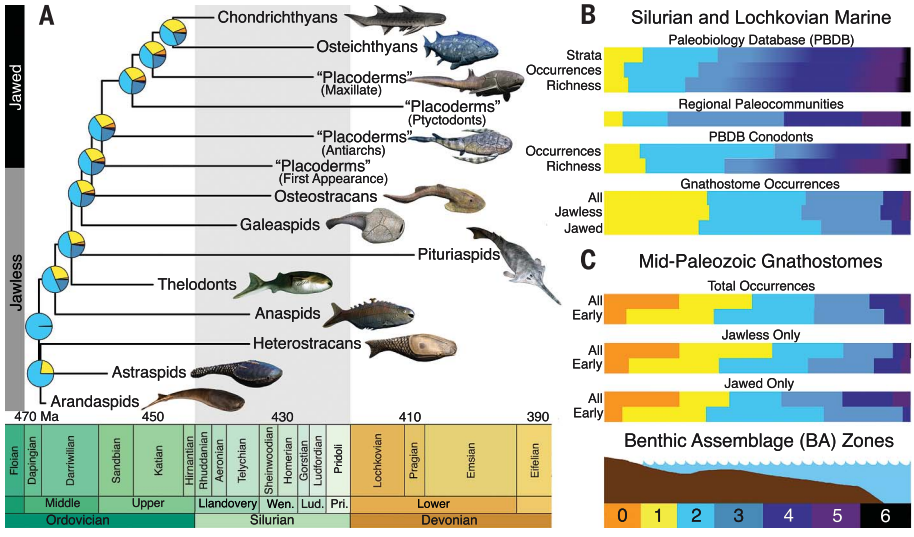


Fig. 1. Mid-Paleozoic vertebrates preferentially originated in shallow marine habitats. (**A**) Intertidal (BA1) to subtidal (BA2 and BA3) ancestral habitats for total-group gnathostome clades (n = 188), assuming placoderm paraphyly and Silurian first occurrence for chondrichthyans. Full results are shown in figs. S2 to S5. Wen., Wenlock; Lud., Ludlow; Pri., Pridoli. (**B**) Silurian and Lochkovian marine distributions for PBDB fossiliferous strata (n = 858), richness (n = 6980), and occurrences (n = 30,004); conodont richness (n = 505) and occurrences (n = 7447); paleocommunities (n = 2401); and gnathostome occurrences (n = 1035) show mid-Paleozoic records peaking on the mid-shelf (BA3 and BA4) with few records in marginal marine settings, in contrast to the shallow-water preferences of early gnathostomes. (**C**) Early and overall occurrences for total-group gnathostomes (n = 2827), jawed fishes (n = 1343), and jawless fishes (n = 1484) show that early occurrences were significantly more concentrated in shallow marine settings (BA1 and BA2) than overall or later occurrences. See data S2 and figs. S1 and S5 to S9.

[1]Department of Earth and Environmental Science, University of Pennsylvania, Philadelphia, PA 19104, USA. [2]Museum of Paleontology, University of Michigan, Ann Arbor, MI 48109, USA. [3]School of Earth and Environmental Sciences, University of Manchester, Manchester M13 9PT, UK. [4]School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham B15 2TT, UK.
***Corresponding author. Email: lsallan@upenn.edu (L.S.); i.j.sansom@bham.ac.uk (I.J.S.)**
†These authors contributed equally to this work.

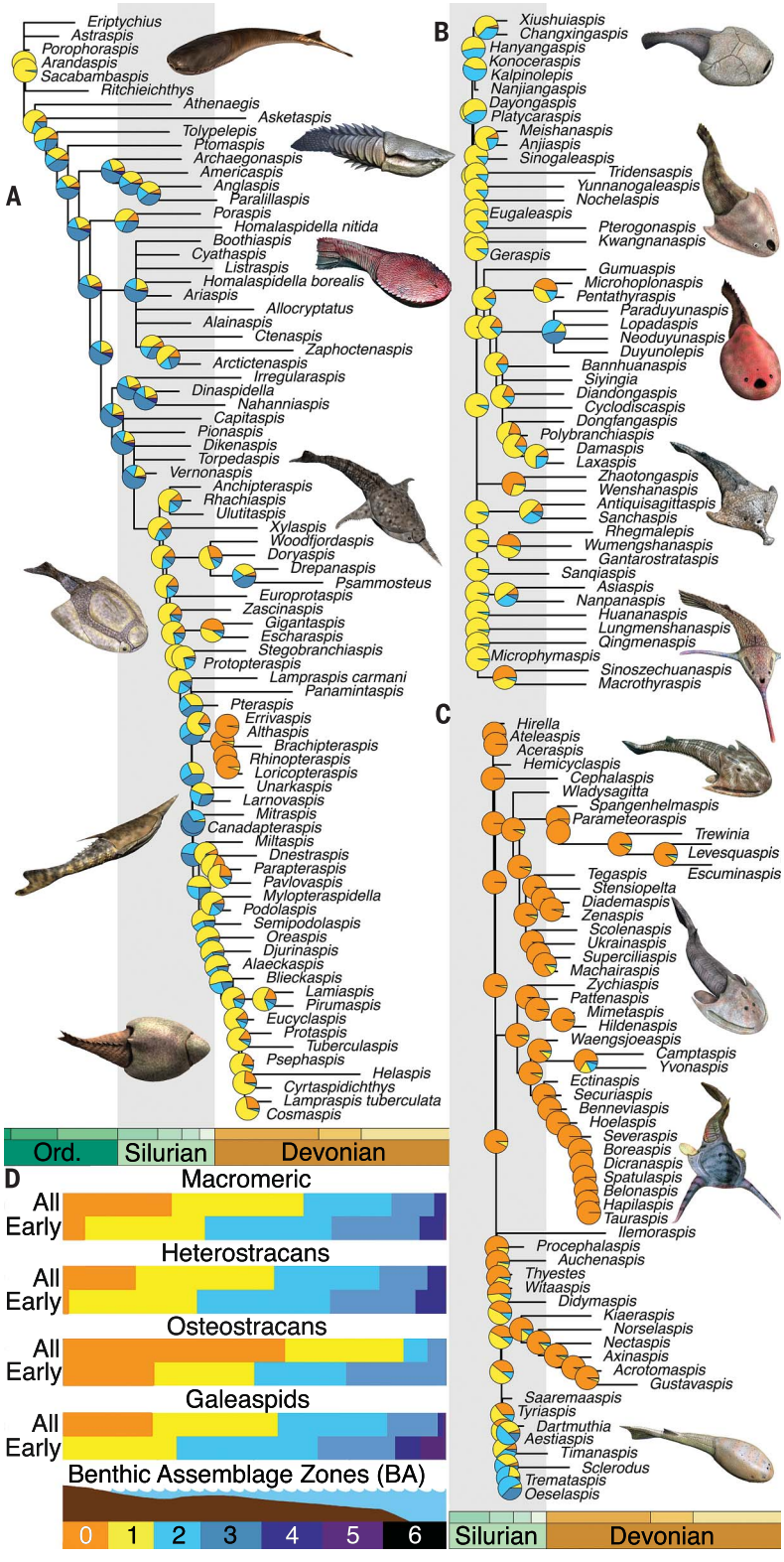


Fig. 2. Macromeric, robust jawless fishes exhibit shallower-water diversification and greater habitat restriction. Ancestral states for (**A**) heterostracans and Ordovician (Ord.) stem-gnathostomes (n = 316), (**B**) galeaspids (n = 112), and (**C**) osteostracans (n = 158) show that macromeric genera preferentially originated in very shallow waters (BA0 to BA2), with the exception of more streamlined forms. Full results are shown in figs S6 to S8. (**D**) Early and overall habitat distributions for macromeric clades (n = 1123), showing significant shifts toward shallower water subsequent to their origination. Full distributions are shown in figs. S19 and S20 and data S1.

(*10*)], with many fewer records in freshwater-marginal marine (BA0 and BA1) and the basin and slope (~BA5 and BA6) zones (Fig. 1B and figs. S11 and S12). PBDB records of occurrences (n = 111,364) or genera (n = 24,211) provide distributions that show even greater clustering on the mid-shelf but are highly correlated with sampled strata (linear regression: r^2 = 0.96, P = 0.0004 and r^2 = 0.94, P = 0.0008, respectively) (fig. S12). Silurian and Lochkovian regional paleocommunities (*10*) are also centered on BA3 and BA4 (Fig. 1B and fig. S13). These records suggest a global, mid-shelf center for sampling and diversity and a null expectation of originations in deep subtidal and reef environments [more so than expected from previous studies focused on reef-bearing facies (*15*)]. This is in stark contrast with shallower gnathostome ancestral habitats (Fig. 1) and is thus unlikely to result from global sampling bias.

To evaluate whether apparent nearshore origination resulted from preservational biases in different habitats, we compared gnathostome distributions to PBDB records for conodonts. Conodonts are the sister group of extant jawless cyclostomes or the vertebrate total-group, largely known from phosphatic oral elements (*4*), which serve as an independent preservational proxy. Conodonts are stratigraphic index fossils and common along the marine depth gradient during the mid-Paleozoic (Fig. 1B and fig. S14). Conodont occurrences (n = 11,915) show a different distribution from other PBDB records (chi-squared P < 0.0001), exhibiting a peak in BA2 and more occurrences in BA5 and BA6 (figs. S14 and S15). Conodont richness (n = 1308) is more clustered around BA3 and BA4, particularly in the Silurian-Lochkovian (n = 505) (Fig. 1B and figs. S14 and S15). This pattern argues against early gnathostome restriction resulting from preservational bias, as does the plurality of vertebrate occurrences in deeper waters from the early Silurian (Fig. 1C and fig. S1).

Jawed and jawless fish distributions are highly clustered in BA0 to BA2 early in clade history (n = 478), in the Silurian and Lochkovian (n = 1035), and over the mid-Paleozoic (n = 2147) (Fig. 1 and figs. S1 and S16 to S18). We recover no significant or strong positive correlations between this gnathostome pattern and other fossil records (linear regression r^2 range: –0.90 to 0.27, P range: 0.41 to 0.9) (Fig. 1B and fig. S16).

Ancestral states show that gnathostomes originated preferentially near shore, even as the diversity of species and body forms increased (Fig. 1A and fig. S2). Early occurrences are significantly different from later records within groups (chi-squared P = < 0.00001) (Fig. 1C and fig. S18); gnathostomes as a whole, as well as jawed and jawless fishes specifically, exhibit greater clustering in shallow marine settings (BA1 and BA2), independent of exact time of first appearance in the mid-Paleozoic (Fig. 1C and fig. S18). Shallow ancestral habitats are always supported by our analyses despite variation in first appearances of jawed fishes (e.g., inclusion of potential Ordovician "chondrichthyan" material) (*15*); placoderm monophyly or paraphyly (*8*); and

even increasing the minimum prior probability of occurrences in all zones to a minimum of 5 or 10% to account for potential of false absence, missing records, or other sampling issues (Fig. 1A, figs. S2 to S5, and table S1). Gnathostomes continued to show a strong tendency to diverge in shallow marine waters long after the invasion of deeper and fresh waters by older lineages, including after the origin of jaws.

Threshold liability values suggest that shifts within the nearshore waters required little evolutionary change and were common, as was invasion of fresh water (Table 1 and Fig. 1C). Dispersal into deeper waters, including the forereef, shelf, and open ocean (BA4 to BA6), was more restricted (Table 1), complicated by a short-term tendency to return to the ancestral shallows [Ornstein-Uhlenbeck deviance information criterion (OU DIC) weight = 1; phylogenetic half-life in Table 1] (*16*). Yet, threshold values also suggest rapid dispersal across the offshore shelf (BA4 and BA5) once lineages managed to depart BA3, even though shifts into open waters (BA6) had much higher requirements (Table 1). However, if sampling probabilities in all bins is increased a priori, shallow-water restriction of early gnathostomes is explained by ever-higher thresholds for continued movement offshore, starting at BA2 (Fig. 1A, figs. S2 to S5, and table S1).

Next, we determined the association between body form and dispersal ability within major groups. Clades were categorized into two body forms: (i) macromeric, which are mostly robust and armored with large bony plates (e.g., heterostracans, osteostracans, and galeaspids) (*17*) (Fig. 2), or (ii) micromeric, which are mostly gracile and either naked or covered in small scales (e.g., thelodonts and anaspids) (*17*) (Fig. 3). These robust or gracile forms can be approximated as having benthic or pelagic and nektonic lifestyles, respectively, given their gross similarity to living fishes (*18*, *19*).

Analysis of all gnathostome early occurrences shows that both micromeric and macromeric forms originated around shallow water (BA2) (Fig. 1A and fig. S2). However, group-level analyses suggest that slight shifts shoreward or seaward preceded the later diversification of these groups. Genus-level diversification of macromeric jawless lineages was centered in the shallows (BA1 and BA2) and fresh water (BA0) throughout their multimillion-year existence (Fig. 2 and figs. S6 to S8, S19, and S20). Later occurrences were significantly more clustered in shallow and freshwater settings than the earliest members of these clades (chi-squared $P < 0.0001$) (Fig. 2C and figs. S19 and S20). Threshold values indicate that moving into deeper waters was more difficult for robust groups than gnathostomes as a whole (Table 1 and tables S1 and S2), and these featured a strong tendency to return to the shallows (OU DIC weight range = 0.99 to 1; phylogenetic half-life in Table 1).

The diversification of micromeric gnathostomes was centered in deeper subtidal waters (BA3) following their origination in BA2 (Figs. 1A and 3 and figs. S9, S10, S21, and S22). Early occurrences of these clades show a significantly greater concentration in BA1 and BA2 than later forms (chi-squared $P < 0.0001$) (fig. S21 and S22). A handful of early Silurian thelodont taxa were already resident in deeper waters (BA3 to BA5) following their Late Ordovician appearances in BA1 and BA2 (fig. S21A). Early dispersal into deeper waters reflects low threshold parameters (Table 1) and may be a general pattern for gracile

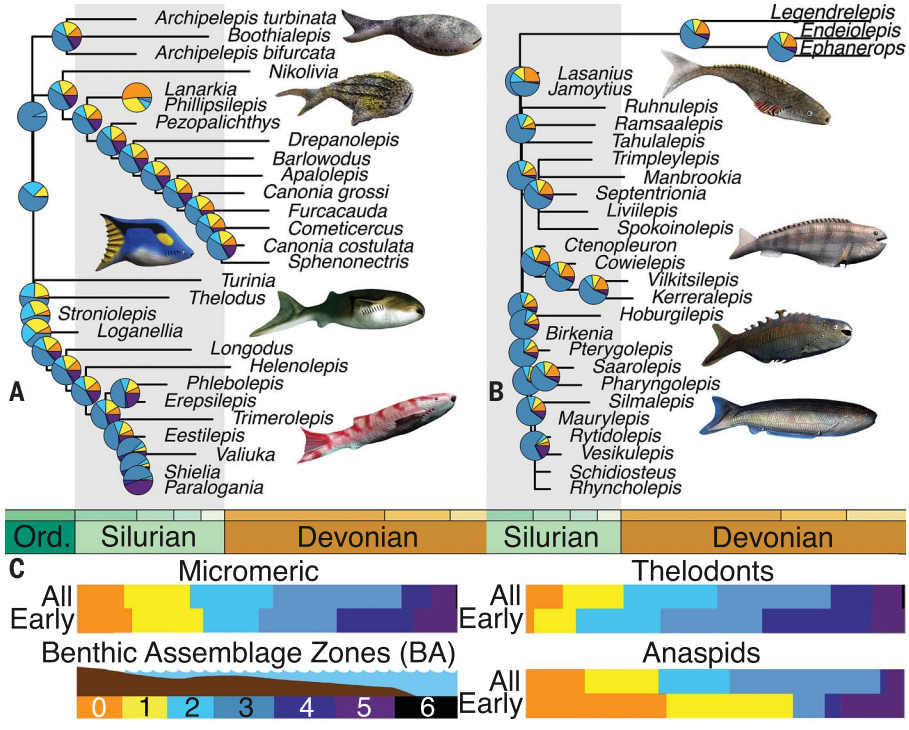


Fig. 3. Micromeric, gracile jawless fishes exhibit deeper-subtidal later diversification and easier dispersal. Ancestral states for (**A**) thelodonts (n = 99) and (**B**) anaspids (n = 100), showing diversification of genera in deeper subtidal waters (BA3) during their evolutionary history. Full results are shown in figs. S9 and S10. (**C**) Early and overall occurrences for micromeric jawless fishes (n = 353) show a rapid shift to deeper waters following nearshore origination. Full distributions are shown in figs. S21 and S22 and data S1.

Table 1. Best-fit model parameters for ancestral habitats. ancThresh (Ancestral Character Estimation Under The Threshold Model Using Bayesian MCMC) (*13*) holds the threshold for exiting BA0 constant at 0 and the value for exiting BA6 as infinity (Inf). Values for parameters (log-likelihood and *alpha*) are means after excluding "burn in." See figs. S2 and S10 and (*29*) for ancestral states. Phylogenetic half-lives are given in million years. mil. gen., million generations.

Clade	Mean threshold liabilities (20 mil. gen., 20% burn-in)							Log-likelihood	*Alpha*	Half-life
	BA0	BA1	BA2	BA3	BA4	BA5	BA6			
Gnathostomes	0	2.09	3.98	6.24	6.81	97.48	Inf	−657.77	0.13	5.33
Heterostracans	0	2.92	3.86	7.74	38.20	200.13	Inf	−979.86	0.12	5.78
Galeaspids	0	3.31	5.91	15.53	83.03	200.55	Inf	−446.63	0.01	138.63
Osteostracans	0	1.13	2.90	26.27	51.66	94.34	Inf	−433.09	0.08	8.66
Anaspids	0	0.19	0.34	1.35	1.40	103.24	Inf	−142.24	1.95	0.36
Thelodonts	0	0.61	0.93	2.05	2.15	110.77	Inf	−220.20	0.59	1.17

clades. Jawed fishes show a significant shift onto reefs and deeper settings in the later Devonian (chi-squared $P < 0.0001$) (Fig. 1C and figs. S1 and S18), after the appearance of most subclades. Robust jawless groups contain exceptions that may prove this rule: A few subclades with fusiform bodies (e.g., trematapsid osteostracans) originated in BA3 and register deeper-water occurrences than their relatives by the mid-Silurian (Fig. 2 and figs. S6 to S8).

Dispersal in multiple directions appears to have been enabled by body-form evolution and did not precede the origin of new phenotypes in new habitats. These shifts affected subsequent survival. Freshwater habitats were marked by the persistence of robust clades such as osteostracans and gracile forms such as anaspids, without further changes to their gross body plan (Figs. 2 and 3). Sometimes, identical deep-water lineages appear short-lived and did not exhibit apparent further diversification, even on reefs (Fig. 1) (*20*). Jawless gnathostomes show a significant shift in distribution (chi-squared $P < 0.00001$) back into the ancestral nearshore habitats and adjacent estuarine areas following a peak in distribution across the depth gradient in the Silurian to Early Devonian (Fig. 1C and figs. S1 and S18). This shift occurred just as jawed fishes moved out of nearshore habitats in the Devonian (Fig. 1A and fig. S18) (*4*, *21*). This pattern is reflected in the greater representation of benthic forms in later marine jawless fishes versus nektonic forms in jawed vertebrates (*22*).

Overall, our results show that the nearshore served as the cradle of early vertebrate taxonomic and gross morphological diversification (Figs. 1 to 3). Specific body forms evolved in coastal waters subsequently favoring expansion into shallower (e.g., macromeric jawless fishes) or deeper areas (e.g., micromeric jawless fishes, jawed fishes). This mirrors observations within living fishes of repeated splits into benthic or pelagic and nektonic forms (*18*, *23*), as well as the gross division of fish phenotype-environment associations (*19*).

A persistent diversification center within the shallows may explain features of the early vertebrate record (*7*, *24*). Ordovician gnathostomes are primarily represented by microfossils restricted to a small subset of nearshore facies (fig. S1) subject to wave action (*11*), despite worldwide distributions (*4*, *7*, *17*, *24*). Ghost lineages for gnathostomes might be caused by environmental endemicity, low abundance, and/or a relative lack of marginal marine strata (figs. S1 and S11 to S13). Alternatively, a relationship between Ordovician diversity and sea level (*6*) might have a common cause in changing shallow habitat area; reduction in such environments would have delayed apparent diversification and increased extinction risk (*6*, *25*, *26*).

Endemicity in coastal waters may have later promoted origination of new clades. Biogeographic patterns suggest that body-form divergence occurred in multiple shallow settings, increasing overall diversity. Micromeric forms occur alongside macromeric astraspids in the Ordovician of Laurentia, whereas robust galeaspids existed alongside gracile chondrichthyans in the early Silurian of Gondwana (*4*–*7*, *15*, *17*, *24*, *27*, *28*). Nektonic body plans developed in these hotspots enabled dispersal across deep early Silurian oceans, away from local competition, leading to further diversification in nearshore settings elsewhere (*1*, *15*, *28*). In contrast, benthic groups showed structured geographic patterns (*27*), moving along coastlines and inshore, perhaps toward nutrient inputs essential to their likely bottom-feeding and filtering lifestyles and away from increased competition. Thus, continuous origination in shallow waters shaped the evolution of vertebrates, at least during their first phase of diversification.

REFERENCES AND NOTES

1. A. J. Boucot, C. Janis, *Paleogeogr. Paleoclimat. Palaeoecol.* **41**, 251–287 (1983).
2. T. C. Chamberlin, *J. Geol.* **8**, 400–412 (1900).
3. G. Carrete Vega, J. J. Wiens, *Proc. Biol. Sci.* **279**, 2323–2329 (2012).
4. M. Friedman, L. C. Sallan, *Palaeontology* **55**, 707–742 (2012).
5. I. J. Sansom, M. M. Smith, M. P. Smith, in *Major Events in Early Vertebrate Evolution*, P. E. Ahlberg, Ed. (Taylor & Francis, 2001), pp. 156–171.
6. R. S. Sansom, E. Randle, P. C. J. Donoghue, *Proc. Biol. Sci.* **282**, 20142245 (2015).
7. N. S. Davies, I. J. Sansom, *Palaios* **24**, 717–722 (2009).
8. M. D. Brazeau, M. Friedman, *Nature* **520**, 490–497 (2015).
9. Materials and methods are available as supplementary materials.
10. A. Boucot, J. Lawson, Eds., *Paleocommunities: A Case Study from the Silurian and Lower Devonian* (Cambridge Univ. Press, 1999).
11. C. E. Brett, A. J. Boucot, B. Jones, *Lethaia* **26**, 25–40 (1993).
12. H. Armstrong, D. A. T. Harper, *Geol. Soc. Am. Spec. Pap.* **505**, 287–300 (2014).
13. L. J. Revell, *Evolution* **68**, 743–759 (2014).
14. Data downloaded from the Paleobiology Database (Paleobiodb.org) 6 to 12 June 2018.
15. P. S. Andreev *et al.*, *J. Vertebr. Paleontol.* **37**, e1245664 (2017).
16. T. F. Hansen, *Evolution* **51**, 1341–1351 (1997).
17. P. Janvier, *Early Vertebrates* (Oxford Univ. Press, 1996).
18. D. Schluter, *The Ecology of Adaptive Radiation* (Oxford Univ. Press, 2000).
19. T. Claverie, P. C. Wainwright, *PLOS ONE* **9**, e112732 (2014).
20. W. Kiessling, C. Simpson, M. Foote, *Science* **327**, 196–198 (2010).
21. P. S. L. Anderson, M. Friedman, M. D. Brazeau, E. J. Rayfield, *Nature* **476**, 206–209 (2011).
22. C. B. Klug *et al.*, *Lethaia* **43**, 465–477 (2010).
23. R. B. Langerhans, *Integr. Comp. Biol.* **48**, 750–768 (2008).
24. M. P. Smith, P. C. J. Donoghue, I. J. Sansom, *Geol. Soc. London Spec. Publ.* **194**, 69–83 (2002).
25. B. Hannisdal, S. E. Peters, *Science* **334**, 1121–1124 (2011).
26. C. Pimiento *et al.*, *Nat. Ecol. Evol.* **1**, 1100–1106 (2017).
27. R. S. Sansom, *Palaeontology* **52**, 1257–1273 (2009).
28. T. Märss, S. Turner, V. Karatajute-Talimaa, in H. P. Schultze Ed., *Handbook of Paleoichthyology* (Verlag Dr. Friedrich Pfeil, 2007), vol. 1B, pp. 1–143.
29. L. Sallan, M. Friedman, R. S. Sansom, C. M. Bird, I. J. Sansom, Data for "The nearshore cradle of early vertebrate diversification" Dryad (2018) 10.5061/dryad.g08m87q.

ACKNOWLEDGMENTS

We thank Z. Min, for assistance with galeaspid occurrences; L. Revell, G. Lloyd, and J. Mitchell for advice on phylogenetic methods; S. Wang for assistance with statistical tests; N. Tamura for providing the reconstructions used in Figs. 1 to 3; and D. Fraser and three anonymous reviewers for comments. **Funding:** This work was supported by the University of Pennsylvania (L.S.), a Palaeontological Association Undergraduate Research Bursary (C.M.B.), and the University of Birmingham (I.J.S). **Author contributions:** L.S. and I.J.S. designed the study, assembled the figures, interpreted results, and drafted the manuscript. L.S., R.S.S., C.M.B., and I.J.S. contributed data. L.S. performed analyses. L.S., M.F., R.S.S., C.M.B., and I.J.S. participated in designing analyses, discussing results, and editing the manuscript. **Competing interests:** None declared. **Data and materials availability:** All data are available in the supplementary materials and at (*29*).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6413/460/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S23
Tables S1 to S3
References (*30*–*61*)
Data S1 to S3

31 October 2017; accepted 2 October 2018
10.1126/science.aar3689

NEUROSCIENCE

PIEZOs mediate neuronal sensing of blood pressure and the baroreceptor reflex

Wei-Zheng Zeng[1], Kara L. Marshall[1], Soohong Min[2], Ihab Daou[1], Mark W. Chapleau[3,4], Francois M. Abboud[3], Stephen D. Liberles[2], Ardem Patapoutian[1]*

Activation of stretch-sensitive baroreceptor neurons exerts acute control over heart rate and blood pressure. Although this homeostatic baroreflex has been described for more than 80 years, the molecular identity of baroreceptor mechanosensitivity remains unknown. We discovered that mechanically activated ion channels PIEZO1 and PIEZO2 are together required for baroreception. Genetic ablation of both *Piezo1* and *Piezo2* in the nodose and petrosal sensory ganglia of mice abolished drug-induced baroreflex and aortic depressor nerve activity. Awake, behaving animals that lack *Piezos* had labile hypertension and increased blood pressure variability, consistent with phenotypes in baroreceptor-denervated animals and humans with baroreflex failure. Optogenetic activation of *Piezo2*-positive sensory afferents was sufficient to initiate baroreflex in mice. These findings suggest that PIEZO1 and PIEZO2 are the long-sought baroreceptor mechanosensors critical for acute blood pressure control.

Blood pressure (BP) is tightly regulated to ensure that the body is prepared to meet varied daily activity demands. Mechanisms that change blood volume control long-term BP regulation. Within seconds and minutes, BP regulation is initiated primarily by baroreceptors, a class of stretch-sensitive neurons within the nodose and petrosal ganglia with peripheral projections in the walls of the aorta and carotid sinus (*1*, *2*). An increase in BP stretches baroreceptor nerve endings to trigger afferent signals that are transmitted to the central nervous system. The consequences of baroreceptor activation are a decrease in heart rate (HR), cardiac output, and vascular resistance that counteract the initial increase in BP (*1*, *2*). Compromised baroreceptor function predicts arrhythmias and premature death in humans with postmyocardial infarction and heart failure (*3*, *4*).

Several ion channels (*5*–*9*) have been suggested to contribute to baroreception. However, substantial residual baroreflex is observed when these channels are ablated, implicating the involvement of other sensory systems. None of the candidate ion channels have been directly activated by mechanical stimuli in heterologous systems, which may lack accessory tethering molecules to form a mechanosensory complex. Furthermore, whether these channels are acting as sensors or play a role downstream of mechanotransduction is not clear. PIEZO1 and PIEZO2 are mechanically activated ion channels that play crucial roles in several mechanotransduction processes (*10*). PIEZO1 is prominently expressed in the cardiovascular system (*11*, *12*), and PIEZO2 is abundant in various populations of sensory neurons (*13*–*15*).

We assessed *Piezo1* and *Piezo2* transcript expression in nodose and petrosal ganglia, where baroreceptor cell bodies are located (*1*). These ganglia are fused with each other and with the jugular ganglion in mice. *Piezo1* and *Piezo2* were highly expressed in the nodose-petrosal-jugular ganglion complex (NPJc) (Fig. 1A). Similar numbers of cells were identified that highly expressed either *Piezo1* or *Piezo2* exclusively (123 and 124 cells with each transcript, respectively, Fig. 1B). A small population of neurons expressed both (43 double *Piezo*-positive cells, or 14.8% of all *Piezo*-expressing cells, n = 6 mice, Fig. 1B).

To test whether *Piezo1* and *Piezo2* are expressed in baroreceptors, we performed retrograde labeling of carotid sensory neurons. We injected fluorescent cholera toxin B (CTB) (*16*) into the carotid sinus beneath the serosal vessel covering. All CTB-positive neurons detected in the NPJc from eight mice were quantified for the presence of *Piezo1* or *Piezo2* transcript (Fig. 1, C to F). Six out of 95 retrogradely labeled cells were *Piezo1*-positive, and eight were *Piezo2*-positive (Fig. 1B). *Piezo*-negative cells were likely chemoreceptors, which abundantly innervate the carotid sinus but do not require mechanosensitivity. None of the 95 CTB-labeled cells

[1]Howard Hughes Medical Institute, Neuroscience Department, Dorris Neuroscience Center, The Scripps Research Institute, La Jolla, CA 92037, USA. [2]Howard Hughes Medical Institute, Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA. [3]Abboud Cardiovascular Research Center, Department of Internal Medicine and Molecular Physiology and Biophysics, Carver College of Medicine, University of Iowa, Iowa City, IA 52242, USA. [4]Veterans Affairs Medical Center, Iowa City, IA 52242, USA.
*Corresponding author. Email: ardem@scripps.edu

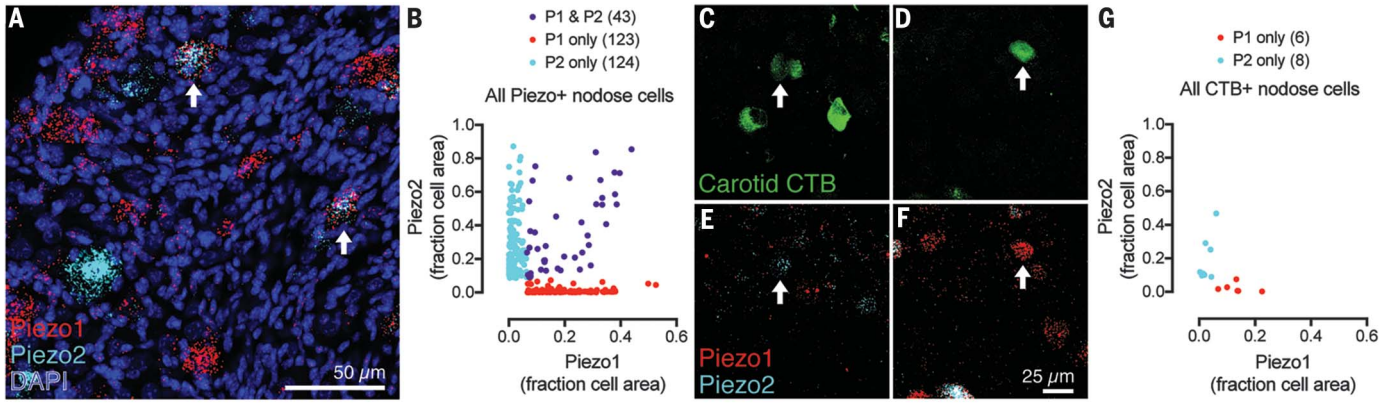


Fig. 1. Expression of *Piezo1* and *Piezo2* transcripts in NPJc. (**A**) Z-projection of NPJc tissue after fluorescent in situ hybridization with probes targeting *Piezo1* (red) and *Piezo2* (cyan). Nuclei are labeled with 4′,6-diamidino-2-phenylindole (DAPI, blue). Arrows mark double *Piezo*-positive cells. (**B**) Quantification of transcript labeling area as a fraction of total cell area (n = 290 cells, six mice). Each dot represents one cell. Numbers in parentheses indicate number of cells. P1, *Piezo1*; P2, *Piezo2*. (**C** to **F**) NPJc cell bodies back-labeled by carotid sinus CTB injections [(C) and (D), green] and *Piezo* transcript [(E) and (F), red and cyan]. Arrows indicate a *Piezo2*-positive cell in (C) and (E) and a *Piezo1*-positive cell in (D) and (F). (**G**) Quantification of *Piezo* transcript labeling area in CTB-positive cells (n = 95 cells, eight mice). *Piezo*-negative cells are not shown.

were double *Piezo*-positive. These data suggest that a subset of neurons that innervate the carotid sinus (which include mechanoreceptors and chemoreceptors) express either *Piezo1* or *Piezo2* (Fig. 1G). We hypothesized that these cells could function as baroreceptors.

We therefore crossed *Piezo* floxed mice to the *Phox2bCre* line, which express Cre recombinase in epibranchial placode-derived ganglia (e.g., nodose and petrosal) but not in neural crest–derived ganglia (jugular, trigeminal, and dorsal root) (*17*). We first analyzed the baroreflex in anesthetized mice in response to phenylephrine (PE). PE induces rapid vasoconstriction (*6*), which elevates BP. Increased BP then triggers baroreceptor activity and induces a reflex decrease in HR. PE-induced baroreflex changes were compared in conditional double-knockout mice (dKO; *Phox2bCre*$^{+}$;*Piezo1*$^{f/f}$*Piezo2*$^{f/f}$) and Cre-negative wild-type littermates (WT). Infusion of PE into the jugular vein produced a dose-dependent and transient increase in systolic BP and a consequent decrease in HR, reflecting baroreflex control (*6*) (Fig. 2A). The PE-induced HR reduction [−29 ± 20 versus −234 ± 24 beats per minute (bpm), $P < 0.001$] and decreased baroreflex sensitivity (−0.6 ± 0.4 versus −5.0 ± 0.5 Δbpm/ΔmmHg, $P < 0.001$) were essentially abolished in the dKO mice (Fig. 2, A to D). PE-induced systolic BP increase in dKO mice was significantly higher than in WT littermates (55.7 ± 3 versus 45.7 ± 6 mmHg, $P < 0.05$) (Fig. 2, A and B). HR response to sodium nitroprusside–induced acute baroreceptor unloading was also absent in dKO mice (fig. S1, A to C). By contrast, *Phox2bCre*$^{+}$;*Piezo1*$^{f/f}$ (P1^{cKO}) and *Phox2bCre*$^{+}$;*Piezo2*$^{f/f}$ (P2^{cKO}) single-knockout mice showed no difference in PE-induced change of baroreflex compared with WT littermates (Fig. 2, B to D). We focused remaining analyses primarily on dKO mice.

We next measured aortic depressor nerve (ADN) activity during a rise of BP induced by PE. We

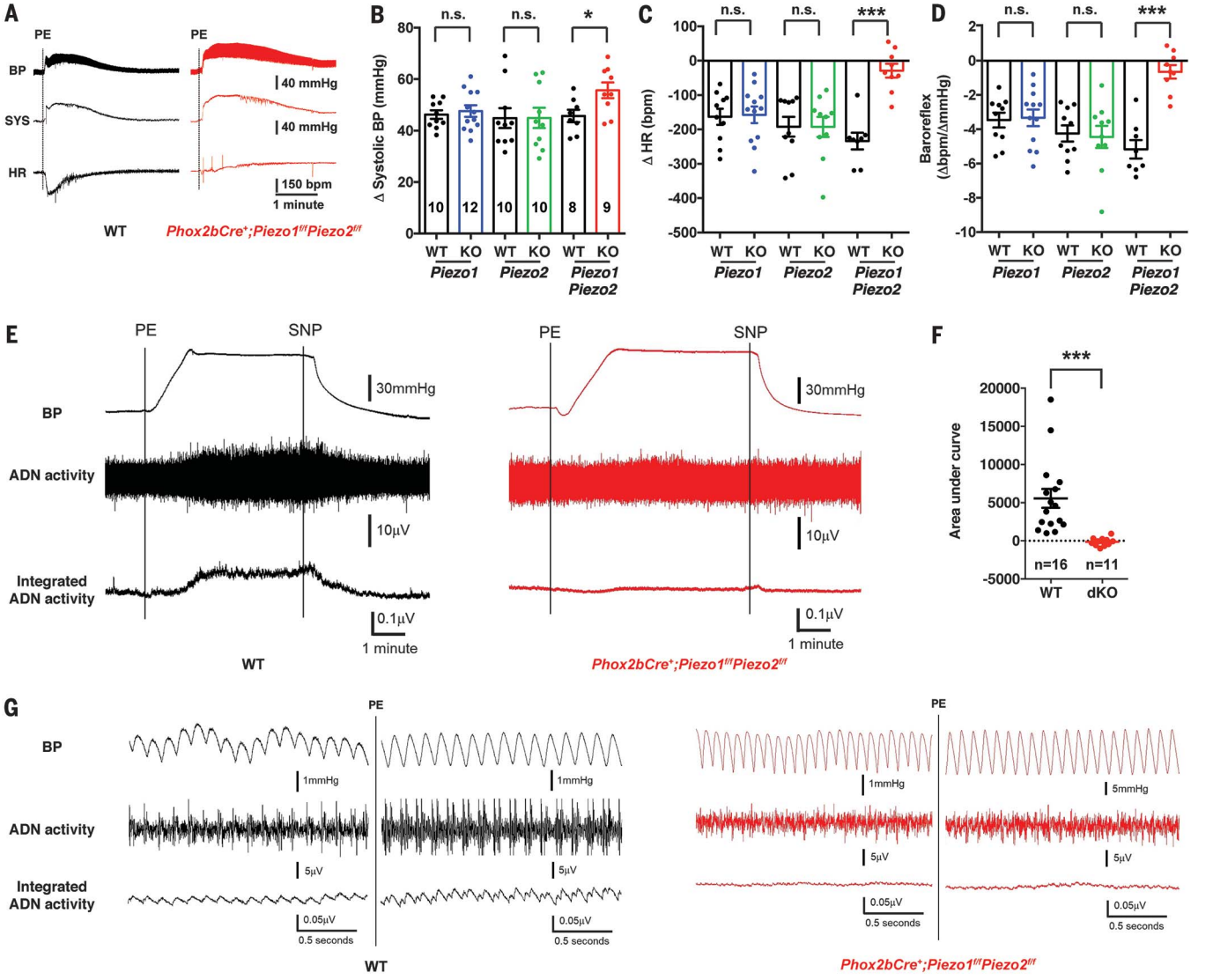


Fig. 2. Baroreflex is abolished in nodose and petrosal ganglia–specific dKO mice. (**A**) Cardiovascular recordings show PE-induced baroreflex in WT mice but no baroreflex in dKO littermates. BP, raw blood pressure signal. SYS, systolic blood pressure derived from raw BP. (**B** to **D**) Changes in systolic BP (B), HR (C), and baroreflex (D) (10 s after intravenous injection of PE) in knockout (KO) mice. Number of animals shown in bars (B) also apply for (C) and (D). Piezo1 KO indicates P1^{cKO} mice; Piezo2 KO indicates P2^{cKO} mice; Piezo1 Piezo2 KO indicates dKO mice. All WT are littermates. (**E**) Traces show BP and ADN activity induced by PE and sodium nitroprusside injection in a WT and a dKO mouse. SNP, sodium nitroprusside. (**F**) Statistical analysis of drug-induced ADN activity in WT ($n = 16$) and dKO ($n = 11$) mice. (**G**) Raw BP and ADN activity example before and after PE injection. Expanded time scale showed bursts of ADN activity in phase with individual arterial pulses in WT. No integrated activity is observed in dKO mice. *$P < 0.05$, ***$P < 0.001$, and n.s. is statistically not significant by unpaired Student's *t* test; data are means ± SEM.

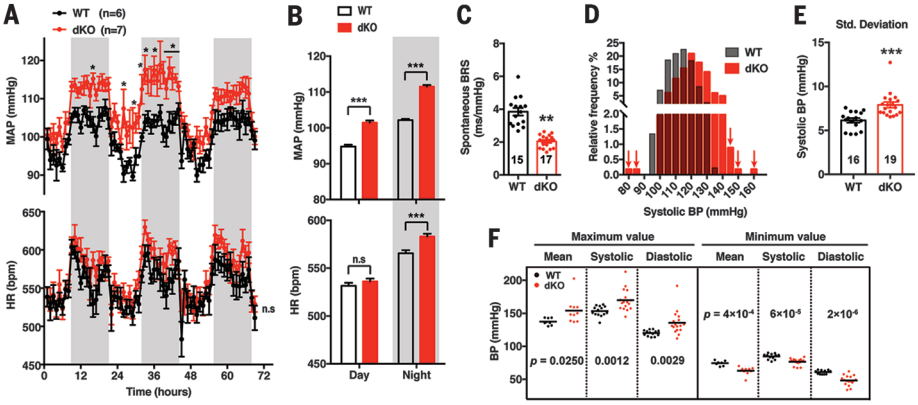


Fig. 3. Increased BP variability in conscious nodose and petrosal ganglia–specific dKO mice. (**A**) Continuous measurements of MAP and HR over 72 hours, binned by hour. The differences between groups were significant during the night (gray shading, two-way analysis of variance, data are means ± SEM). (**B**) Average MAP and HR during the day (6 a.m. to 6 p.m.) and night (6 p.m. to 6 a.m.). (**C**) sBRS, expressed as change in PI (ms) per change in systolic BP (mmHg), was significantly reduced in dKO mice (n = 17 mice) compared with WT mice (n = 15 mice). (**D**) Frequency distribution histogram of the systolic BP over 72 hours. Red arrows indicate wider distribution of BP in dKO mice. (**E**) BP variability reported as standard deviation from the 72-hour period. (**F**) Maximum and minimum BP values. P values are indicated in the bars. *P < 0.05, **P < 0.01, ***P < 0.001, and n.s. is statistically not significant. Unpaired Student's t test was used unless indicated otherwise; data are means ± SEM. n = 7 to 17 mice.

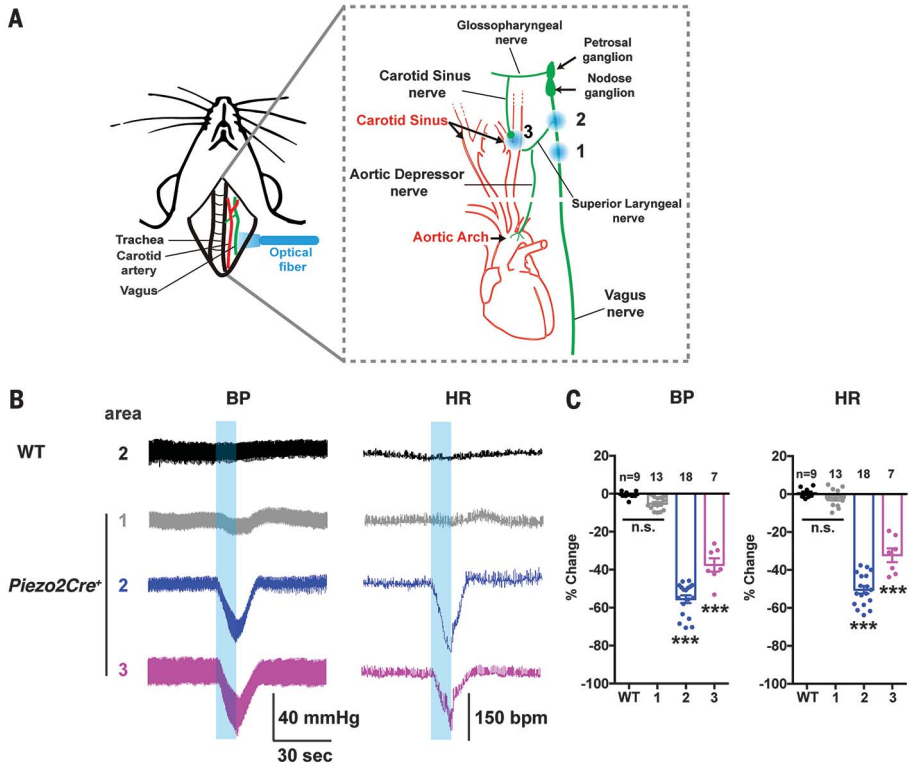


Fig. 4. *Piezo2*-positive sensory neurons acutely control blood pressure. (**A**) Schematic depiction of the optogenetic strategy. The carotid sinus and vagus nerves were illuminated to activate ChR2-expressing *Piezo2*-sensory neurons. The optical fiber was placed on area 1, vagus nerve trunk; area 2, superior laryngeal branch; and area 3, carotid sinus. (**B**) Traces of cardiovascular effects after focal vagus nerve illumination (light blue shading) in anesthetized *Piezo2Cre*$^{-}$;*ChR2-eYFP* (WT, black trace) and *Piezo2Cre*$^{+}$;*ChR2-eYFP* mice (*Piezo2Cre*$^{+}$, gray, blue, and pink traces). Numbers on the left (1 to 3) correspond to areas in (A). Blood pressure was measured by a pressure transducer cannulated in the left carotid artery. BP, carotid arterial pressure. (**C**) Light-induced changes in BP and HR were calculated over the 10 s (n = 7 to 18 mice, as indicated; ***P < 0.001, and n.s. is statistically not significant by unpaired Student's t test; data are means ± SEM).

observed a lack of drug-induced ADN activity in dKO mice compared to WT mice (−131.9 ± 163.9 versus 5558 ± 1234 normalized area under curve of integrated ADN activity, P < 0.001; Fig. 2, E to G). The dKO mice had no appreciable responses during both phasic and tonic phases of PE-induced ADN activity (fig. S1, D and E). This is not due to gross anatomical deficits, because we observed comparable baroreceptor ending densities within the aortic arch of dKO and WT mice (fig. S2).

Impaired baroreceptor function leads to dysregulation of BP, including volatile hypertension and increased BP variability in humans (*18*–*20*). We examined daily BP variability in freely moving, conscious mice using a telemetric sensor (*6*). The dKO mice showed significantly increased mean arterial pressure (MAP) during their active time (gray shading, 6 p.m. to 6 a.m.) compared with WT littermates (112 ± 0.4 versus 95 ± 0.5 mmHg, P < 0.001) (Fig. 3, A and B, and figs. S3, A and B, and S4). The HR of dKO mice was slightly increased during active times compared with that of WT mice (583 ± 3 versus 566 ± 3 bpm, P < 0.001), whereas the HR remained unchanged during inactive times (6 a.m. to 6 p.m., 532 ± 3 versus 536 ± 3 bpm, not significant) (Fig. 3B). No difference in locomotor activity was observed between dKO and WT mice (fig. S3C), ruling out the possibility that activity caused the increased BP and HR in dKO mice.

We scanned telemetry data for spontaneous changes in systolic BP and pulse interval (PI) consistent with a baroreflex relationship. This method (sequence technique) is used to noninvasively assess baroreflex function (*21*, *22*). The spontaneous baroreflex sensitivity (sBRS) is defined as the slope of changes in systolic BP versus PI from 1 hour of recording. sBRS was severely reduced in dKO mice (2.0 ± 0.1 versus 3.8 ± 0.2 ms/mmHg for WT, P < 0.001, Fig. 3C). Sinoaortic baroreceptor denervated mice also show residual sBRS (*22*), and this may be due to compensation from other sensory systems.

We compared the BP variability of WT and dKO mice. The systolic BP values of dKO mice were distributed in a broader range than those of WT littermates (Fig. 3D). Variability was greatly enhanced in dKO mice (7.9 ± 0.3 versus 6.1 ± 0.3 mmHg in WT, P < 0.001, Fig. 3E). We quantified the range of BP variability of mean, systolic, and diastolic BP within each group in a 72-hour period. Maximum values of BP from dKO mice were significantly higher than those from WT littermates, whereas minimum values were significantly lower (Fig. 3F). Lastly, homovanillic acid concentrations in dKO mouse urine were significantly higher than those in WT urine (13.9 ± 0.05 versus 12.1 ± 0.06 μg/ml, fig. S3D), suggesting an increase in hormone norepinephrine concentration, as in human baroreflex failure patients (*18*). There were no significant BP variability and sBRS differences in P1^{cKO} (fig. S5) and P2^{cKO} (fig. S6) single-knockout mice compared with WT littermates. P2^{cKO} mice showed a subtle hypotensive BP distribution (fig. S6).

We next investigated whether stimulating *Piezo2*-positive neurons can induce the baroreflex in adult mice. We crossed *Piezo2^{GFP}-IRES-Cre* (*Piezo2Cre*) knockin mice with Cre-dependent *channelrhodopsin-2* (*ChR2*) reporter mice to generate *Piezo2Cre^{+};ChR2-eYFP* mice (*13*) and recorded the cardiovascular response to activating different regions of *Piezo2*-positive vagal sensory nerves by optogenetics (Fig. 4A; *eYFP*, enhanced yellow fluorescent protein gene). We did not observe cardiovascular changes during long optogenetic stimulation (5-ms pulses, 50 Hz, 10 s) of the vagal nerve trunk (area 1 in Fig. 4A; BP, −5.3 ± 1.0%, and HR, −2.0 ± 1.0%; not significant) (Fig. 4, A to C). Next, we focused on specifically activating baroreceptor afferents. For aortic baroreceptors, we stimulated the superior laryngeal nerve branch, which carries afferent inputs from the ADN (area 2 in Fig. 4A). For carotid baroreceptors, we exposed the carotid sinus region and directly stimulated the local nerve terminals (area 3 in Fig. 4A). Light stimulations at both locations caused an immediate decrease in both BP and HR (area 2: BP, −55.6 ± 2.0%, and HR, −50.5 ± 2.0%; area 3: BP, −37.5 ± 3.5%, and HR, −32.3 ± 3.7%; P <0.001) compared with the unstimulated baseline (Fig. 4, B and C). A prominent consequence of baroreceptor activation is rapid inhibition of efferent sympathetic activity (*1*). We found that light-induced decrease in HR was markedly attenuated after administration of the β-adrenergic receptor–blocker propranolol, indicating that the reflex bradycardia was mediated primarily by inhibition of cardiac sympathetic nerve activity (fig. S7). *Piezo2Cre^{-};ChR2-eYFP* mice (WT) did not show any changes during optogenetic stimulation in all three regions (BP, −0.6 ± 0.6%, and HR, 0.4 ± 0.8%; not significant; Fig. 4, B and C).

This study demonstrates that the mechanically activated ion channels PIEZO1 and PIEZO2 are together required for arterial baroreceptor activity and function. Baroreflex is critical to maintain short-term BP homeostasis in mammals. The long-term changes observed in HR and BP that accompany baroreflex failure are complex. Acute elimination of baroreceptor function (e.g., sinoaortic denervation) causes immediate, large increases in BP and HR (*23*, *24*). Over time, the mean BP decreases but remains labile hypertensive, and BP variability is markedly increased and persists (*18–20*, *24*, *25*). We observed a significant increase in MAP during the active period of the *Piezo* dKO mice that falls just under the designation for hypertension (*26*), and dKO mice also developed increased blood pressure variability. These data show that losing PIEZO1 and PIEZO2 function recapitulates the phenotype observed in animal models (*24*, *25*) and humans with baroreflex failure (*18–20*). However, we cannot exclude the possibility that sensory mechanisms beyond the baroreceptors within the vagus contribute to the observed increased blood pressure.

REFERENCES AND NOTES

1. E. A. Wehrwein, M. J. Joyner, *Handb. Clin. Neurol.* **117**, 89–102 (2013).
2. H. R. Kirchheim, *Physiol. Rev.* **56**, 100–177 (1976).
3. A. Mortara *et al.*, *Circulation* **96**, 3450–3458 (1997).
4. M. T. La Rovere, *Lancet* **351**, 478–484 (1998).
5. H. A. Drummond, M. P. Price, M. J. Welsh, F. M. Abboud, *Neuron* **21**, 1435–1441 (1998).
6. Y. Lu *et al.*, *Neuron* **64**, 885–897 (2009).
7. A. T. Chesler *et al.*, *N. Engl. J. Med.* **375**, 1355–1364 (2016).
8. O. C. Lau, B. Shen, C. O. Wong, X. Yao, *Nat. Commun.* **9**, 1244 (2018).
9. P. Thakore, S. D. Brain, D. J. Beech, *Nat. Commun.* **9**, 1245 (2018).
10. S. E. Murthy, A. E. Dubin, A. Patapoutian, *Nat. Rev. Mol. Cell Biol.* **18**, 771–783 (2017).
11. S. Wang *et al.*, *J. Clin. Invest.* **126**, 4527–4536 (2016).
12. K. Retailleau *et al.*, *Cell Reports* **13**, 1161–1171 (2015).
13. K. Nonomura *et al.*, *Nature* **541**, 176–181 (2017).
14. S. H. Woo *et al.*, *Nat. Neurosci.* **18**, 1756–1762 (2015).
15. S. S. Ranade *et al.*, *Nature* **516**, 121–125 (2014).
16. K. L. Marshall *et al.*, *Cell Reports* **11**, 851–858 (2015).
17. A. Pattyn, X. Morin, H. Cremer, C. Goridis, J. F. Brunet, *Development* **124**, 4065–4075 (1997).
18. D. Robertson *et al.*, *N. Engl. J. Med.* **329**, 1449–1455 (1993).
19. K. Heusser, J. Tank, F. C. Luft, J. Jordan, *Hypertension* **45**, 834–839 (2005).
20. T. Ketch, I. Biaggioni, R. Robertson, D. Robertson, *Circulation* **105**, 2518–2523 (2002).
21. J. Oosting, H. A. Struijker-Boudier, B. J. Janssen, *J. Hypertens.* **15**, 391–399 (1997).
22. P. Martinka *et al.*, *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **288**, R767–R776 (2005).
23. I. C. Wenker *et al.*, *J. Neurosci.* **37**, 4565–4583 (2017).
24. F. L. Rodrigues, M. de Oliveira, H. C. Salgado, R. Fazan Jr., *Exp. Physiol.* **96**, 853–862 (2011).
25. C. S. Ito, A. M. Scher, *Circ. Res.* **48**, 576–591 (1981).
26. P. A. James *et al.*, *JAMA* **311**, 507–520 (2014).

ACKNOWLEDGMENTS

We thank D. Morgan, S. Ma, and K. Nonomura for assistance and D. Ginty for the suggestion to assess the role of PIEZO2 in baroreceptors. **Funding:** This work was supported by NIH grants R01 DE022358 and R35 NS105067 to A.P. W.-Z.Z. was supported by a postdoctoral fellowship from the George Hewitt Foundation for Medical Research. S.D.L. was supported by NIH grants DP1 AT009497 and OT2 OD023848. M.W.C and F.M.A were supported by NIH grant P01 HL14388. A.P. and S.D.L. are investigators of the Howard Hughes Medical Institute. **Author contributions:** W.-Z.Z., K.L.M., and A.P. designed experiments and wrote the paper. K.L.M. performed in situ hybridization and baroreceptor innervation analysis. W.-Z.Z. performed drug-induced baroreflex assessment, telemetry sensor implantation, BP variability, and sBRS analysis. W.-Z.Z. and I.D. performed optogenetics experiments. S.M. performed ADN activity recordings in the S.D.L. laboratory. M.W.C and F.M.A advised and trained W.-Z.Z., contributed to technical approaches, and edited the manuscript. **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

www.sciencemag.org/content/362/6413/464/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S7
References (*27*, *28*)

29 June 2018; accepted 7 September 2018
10.1126/science.aau6324

2019 MRS® SPRING MEETING & EXHIBIT

April 22–26, 2019 | Phoenix, Arizona

CALL FOR PAPERS

Abstract Submission Opens
September 28, 2018

Abstract Submission Closes
October 31, 2018

Spring Meeting registrations include MRS Membership July 1, 2019 – June 30, 2020

BROADER IMPACT

BI01 High Impact Practice—Increasing Ethnic and Gender Diversification in Engineering Education

CHARACTERIZATION, PROCESSING AND THEORY

CP01 Advances in *In Situ* Experimentation Techniques Enabling Novel and Extreme Materials/Nanocomposite Design
CP02 Design and *In Situ* TEM Characterization of Self-Assembling Colloidal Nanosystems
CP03 Advances in *In Situ* Techniques for Diagnostics and Synthetic Design of Energy Materials
CP04 Interfacial Science and Engineering—Mechanics, Thermodynamics, Kinetics and Chemistry
CP05 Materials Evolution in Dry Friction—Microstructural, Chemical and Environmental Effects
CP06 Smart Materials for Multifunctional Devices and Interfaces
CP07 From Mechanical Metamaterials to Programmable Materials
CP08 Additive Manufacturing of Metals
CP09 Mathematical Aspects of Materials Science—Modeling, Analysis and Computations

ELECTRONICS AND PHOTONICS

Soft Organic and Bimolecular Electronics

EP01 Liquid Crystalline Properties, Self-Assembly and Molecular Order in Organic Semiconductors
EP02 Photonic Materials and Devices for Biointerfaces
EP03 Materials Strategies and Device Fabrication for Biofriendly Electronics
EP04 Soft and Stretchable Electronics—From Fundamentals to Applications
EP05 Engineered Functional Multicellular Circuits, Devices and Systems
EP06 Organic Electronics—Materials and Devices

Semiconductor Devices, Interconnects, Plasmonic and Thermoelectric Materials

EP07 Next-Generation Interconnects—Materials, Processes and Integration
EP08 Phase-Change Materials for Memories, Photonics, Neuromorphic and Emerging Application
EP09 Devices and Materials to Extend the CMOS Roadmap for Logic and Memory Applications
EP10 Heterovalent Integration of Semiconductors and Applications to Optical Devices
EP11 Hybrid Materials and Devices for Enhanced Light-Matter Interactions
EP12 Emerging Materials for Plasmonics, Metamaterials and Metasurfaces
EP13 Thermoelectrics—Materials, Methods and Devices

ENERGY AND SUSTAINABILITY

Energy Storage

ES01 Organic Materials in Electrochemical Energy Storage
ES02 Next-Generation Intercalation Batteries
ES03 Electrochemical Energy Materials Under Extreme Conditions
ES04 Solid-State Electrochemical Energy Storage

Catalysis, Alternative Energy and Fuels

ES05 Cooperative Catalysis for Energy and Environmental Applications
ES06 Atomic-Level Understanding of Materials in Fuel Cells and Electrolyzers
ES07 New Carbon for Energy—Materials, Chemistry and Applications
ES08 Materials Challenges in Surfaces and Coatings for Solar Thermal Technologies
ES10 Rational Designed Hierarchical Nanostructures for Photocatalytic System
ES11 Advanced Low Temperature Water-Splitting for Renewable Hydrogen Production via Electrochemical and Photoelectrochemical Processes
ES12 Redox-Active Oxides for Creating Renewable and Sustainable Energy Carriers

Water-Energy Materials and Sustainability

ES09 Advanced Materials for the Water-Energy Nexus
ES13 Materials Selection and Design—A Tool to Enable Sustainable Materials Development and a Reduced Materials Footprint
ES14 Materials Circular Economy for Urban Sustainability

Photovoltaics and Energy Harvesting

ES15 Fundamental Understanding of the Multifaceted Optoelectronic Properties of Halide Perovskites
ES16 Perovskite Photovoltaics and Optoelectronics
ES17 Perovskite-Based Light-Emission and Frontier Phenomena—Single Crystals, Thin Films and Nanocrystals
ES18 Frontiers in Organic Photovoltaics
ES19 Excitonic Materials and Quantum Dots for Energy Conversion
ES20 Thin-Film Chalcogenide Semiconductor Photovoltaics
ES21 Nanogenerators and Piezotronics

QUANTUM AND NANOMATERIALS

QN01 2D Layered Materials Beyond Graphene—Theory, Discovery and Design
QN02 Defects, Electronic and Magnetic Properties in Advanced 2D Materials Beyond Graphene
QN03 2D Materials—Tunable Physical Properties, Heterostructures and Device Applications
QN04 Nanoscale Heat Transport—Fundamentals
QN05 Emerging Thermal Materials—From Nanoscale to Multiscale Thermal Transport, Energy Conversion, Storage and Thermal Management
QN06 Emerging Materials for Quantum Information
QN07 Emergent Phenomena in Oxide Quantum Materials
QN08 Colloidal Nanoparticles—From Synthesis to Applications

SOFT MATERIALS AND BIOMATERIALS

SM01 Materials for Biological and Medical Applications
SM02 Progress in Supramolecular Nanotheranostics
SM03 Growing Next-Generation Materials with Synthetic Biology
SM04 Translational Materials in Medicine—Prosthetics, Sensors and Smart Scaffolds
SM05 Supramolecular Biomaterials for Regenerative Medicine and Drug Delivery
SM06 Nano- and Microgels
SM07 Bioinspired Materials—From Basic Discovery to Biomimicry

www.mrs.org/spring2019

Meeting Chairs

Yuping Bao The University of Alabama
Bruce Dunn University of California, Los Angeles
Subodh Mhaisalkar Nanyang Technological University
Ruth Schwaiger Karlsruhe Institute of Technology—Institute for Applied Materials
Subhash L. Shinde University of Notre Dame

Don't Miss These Future MRS Meetings!

2019 MRS Fall Meeting & Exhibit
December 1–6, 2019, Boston, Massachusetts

2020 MRS Spring Meeting & Exhibit
April 13–17, 2020, Phoenix, Arizona

MRS MATERIALS RESEARCH SOCIETY®
Advancing materials. Improving the quality of life.
506 Keystone Drive • Warrendale, PA 15086-7573
Tel 724.779.3003 • Fax 724.779.8313 • info@mrs.org • www.mrs.org

BIOLOGY ▪ EARTH ▪ ENVIRONMENT ▪ LIFE ▪ INTERDISCIPLINARY ▪ PHYSICAL ▪ MATERIAL ▪ SOCIAL SCIENCES

Science Advances | AAAS

OPEN ACCESS, DIGITAL, AND FREE TO ALL READERS

OPEN ACCESS

Pushing the Boundaries of Knowledge

As AAAS's first multidisciplinary, open access journal, *Science Advances* publishes research that reflects the selectivity of high impact, innovative research you expect from the *Science* family of journals, published in an open access format to serve a vast and growing global audience. Check out the latest findings or learn how to submit your research: **ScienceAdvances.org**

Anti-Ranibizumab Antibodies

Bio-Rad offers a range of recombinant monoclonal anti-ranibizumab antibodies that are highly specific for the monoclonal antibody drug ranibizumab (Lucentis) or the complex of ranibizumab with its target, vascular endothelial growth factor A (VEGF-A). Ranibizumab is used for the treatment of wet age-related macular degeneration (AMD), a common form of age-related vision loss, and for macular edema and diabetic retinopathy. The anti-ranibizumab antibodies are inhibitory, noninhibitory, and drug-target-complex specific antibodies, and are designed for use in pharmacokinetic (PK) and immunogenicity assays for ranibizumab and biosimilars. Bio-Rad's anti-idiotypic antibodies are generated using Human Combinatorial Antibody Library (HuCAL) and CysDisplay, a proprietary method of phage display with guided selection methods to obtain highly targeted reagents. The recombinant production method also ensures a consistent, secure supply. The anti-ranibizumab antibodies are approved for in vitro research and commercial applications of in vitro testing services that support preclinical and clinical drug and biosimilar development and patient monitoring.

Bio-Rad
For info: 800-424-6723
www.bio-rad-antibodies.com/anti-ranibizumab-antibody-lucentis.html

Reaction Monitoring System NMR Tube

The Reaction Monitoring System NMR Tube from Wilmad-LabGlass permits researchers to run reaction experiments without the need for costly instruments or equipment. This two-chamber, borosilicate glass tube is a valuable tool for use in kinetic studies and reaction-endpoint determinations. The system features an inner chamber with a Teflon tip that, when secured in the outer chamber, creates a positive seal separating solutions until the user is ready to mix. The in situ mixing of solutions allows for clear before-and-after reaction spectra. With the use of a benchtop nuclear magnetic resonance (NMR) spectrometer where the upper portion of the sample tube is accessible, researchers may also be able to capture initial kinetic data points. Available in 7-in. and 8-in. tube lengths, the Reaction Monitoring System comes with an extra black phenolic cap for postmix storage of solutions.

Wilmad-LabGlass
For info: 800-220-5171
www.wilmad-labglass.com

Evaporators

Genevac's HT Series 3 evaporators with the Inert Gas Purge (IGP) option enable safe operation even when removing highly explosive solvents such as diethyl ether. To eliminate these risks, the IGP system purges the air from the evaporators by using an inert gas prior to commencing the evaporation process. The IGP option can also be useful where an inert atmosphere is required to help protect hygroscopic or oxygen-sensitive samples at the end of the evaporation or concentration process; the inert-gas atmosphere is maintained until the user opens the system to retrieve their samples. Incorporating a high-performance vacuum pump, the latest touchscreen technology, and a sleek ergonomic design, the HT Series 3 makes optimizing your evaporation processes effortless, while accommodating a wide range of sample formats, including microtubes, microplates, and flasks.

Genevac
For info: 845-255-5000
www.spscientific.com/ht3

Breath Biopsy Kits

Breath Biopsy Kits from Owlstone Medical allow academic, clinical, and pharmaceutical researchers to quickly discover and validate breath-based biomarkers in early detection and precision medicine research activities. VOC (volatile organic compound) biomarkers present in breath can be analyzed with high sensitivity and selectivity in Owlstone Medical's Breath Biopsy Services Laboratory, or in your own laboratory. The kits include a Breath Biopsy Cartridge to collect, stabilize, and enrich VOCs in breath, plus a disposable Breath Biopsy Mask. All kit components are manufactured and quality-checked to exacting standards. ReCIVA Breath Samplers are available to purchase or lease. By sending samples to our breath biopsy lab for analysis, scientists can benefit from the world's largest breath biopsy digital biobank, to further explore biomarker classifiers and gain valuable insights into potential performance in larger patient populations.

Owlstone Medical
For info: +44-(0)-1223-428200
www.owlstonemedical.com

Isogenic Cell Lines for Cancer Research

The ISOMAX range of isogenic engineered cell-line products from AMS Biotechnology (AMSBIO) feature diverse oncogenic driver mutations in various pathways, including the MAPK and mTOR interactive signaling pathways. Over 150-plus point mutations engineered across more than 15 genes are available, such as MAPK pathway line diverse mutations in *EGFR*, *KRAS*, and *BRAF* genes. All cell lines are available with genetically matched "normal cells" to provide a toolset for enhancing our understanding of cancer biology, or for high-throughput screening to identify lead/target compounds for cancer therapeutics. Using CRISPR/Cas9 technology, these cells lines are free from genomic footprints that can impact scientific results, and do not require selection markers. All cell lines are expanded from single-cell clones and validated at the genomic level by Sanger sequencing. By generating clonal populations of mutant cell lines, AMSBIO can ensure the highest quality and homogeneity of the engineered cells. Alongside this comprehensive library of cell lines, AMSBIO offers custom cell-line engineering services.

AMS Biotechnology
For info: +44-(0)-1235-828200
www.amsbio.com/isogenic-cell-lines.aspx

High-Flow Sterile Gas Filter

Amazon Filters SupaPore THB filter cartridges are rated to 0.02 µm to ensure effective removal of contaminants from gases that come into product contact during biopharmaceutical or pharmaceutical sterile-packaging processes. Designed to meet U.S. FDA 21 CFR (Code of Federal Regulations) Title 21 requirements, these cartridges provide a barrier that guarantees the sterility of the process. The filter removes viruses and microorganisms from the gas as it flows either in the facility's compressed-gas main or in and out of the product storage tank. Manufactured using a highly hydrophobic polytetrafluoroethylene (PTFE) membrane, the SupaPore THB filter cartridge offers exceptionally high gas-flow rates at low-pressure differentials, thereby minimizing operating costs. All cartridges are thermally bonded and adhesive-free. Each unit is preflushed with ultrapure water and integrity-tested before final assembly. The cartridges are proven to deliver trouble-free operation even after many steam-sterilization cleaning cycles.

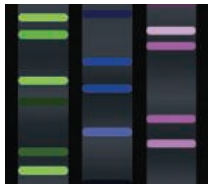
Amazon Filters
For info: +44-(0)-1276-670600
www.amazonfilters.com

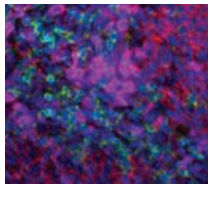
Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

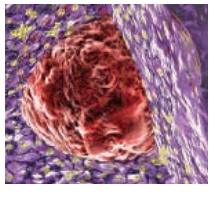
Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

want new technologies?

antibodies
apoptosis
biomarkers
cancer
cytometry
data
diseases
DNA
epigenetics
genomics
immunotherapies
medicine
microbiomics
microfluidics
microscopy
neuroscience
proteomics
sequencing
toxicology
transcriptomics







watch our webinars

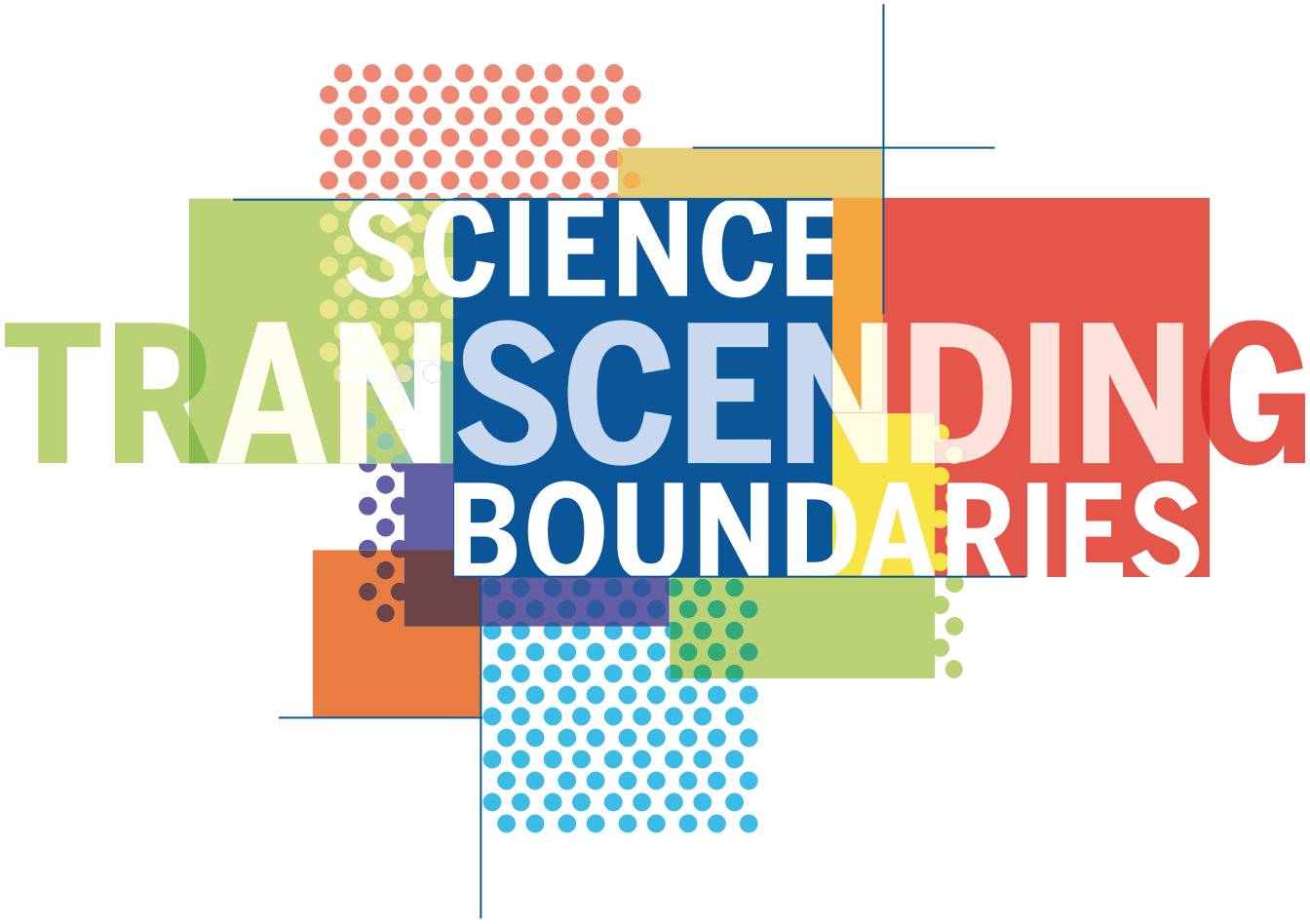
Learn about the latest breakthroughs, new technologies, and ground-breaking research in a variety of fields. Our expert speakers explain their quality research to you and answer questions submitted by live viewers.

VIEW NOW!
webinar.sciencemag.org

Science
AAAS

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

@SciMagWebinars



BE PART OF THE CONVERSATION

Artificial Intelligence | Astrophysics | Genomics | Sustainability
Science Diplomacy | Food Security | Digital Privacy | Infection Control
Public Engagement | Climate Change | Diversity | Big Data

AAAS
ANNUAL MEETING
Washington, DC | Feb. 14–17, 2019

SEE INSIDE FOR DETAILS:

Topical Lectures | Milestone Celebrations
Scientific Sessions | Career Workshops
Flash Talk Program | Registration Rates

President's Address

Dear Colleagues:

Many lines are drawn in today's world: within our communities, within global society, and within science itself. The 2019 AAAS Annual Meeting theme, Science Transcending Boundaries, considers how science can bring together people, ideas, and solutions from across real and artificial borders, disciplines, sectors, ideologies, and traditions.

On behalf of the AAAS Board of Directors, I urge you to join us in Washington, DC, February 14-17, 2019, where this theme will be explored through interdisciplinary scientific sessions, renowned speakers, and in-depth discussions. The AAAS Annual Meeting is the most widely reported global science gathering and the premier event at which you can network with future collaborators across disciplines. We look forward to seeing you in Washington.

Margaret Hamburg
AAAS President
Foreign Secretary, National Academy of Medicine

Join us in Washington, DC

Learn about the many ways that scientific research and diplomacy overcome disciplinary and geographic limitations:

- Plenary and topical sessions by accomplished speakers
- Celebrations of scientific milestones
- 120+ scientific sessions in 11 disciplinary tracks covering exciting advances in research and policy
- 20+ flash talk sessions on current scientific projects and initiatives
- Opportunities to network with colleagues and attend workshops on career guidance
- E-poster presentations highlighting science research by students and professionals

CONNECT WITH US!

@AAASmeetings
#AAASmtg

facebook.com/AAAS.Science

aaas.org/meetings

REPORTERS

The AAAS Annual Meeting Newsroom will be hosted on EurekAlert! at eurekalert.org/aaasnewsroom

Sponsors As of September 2018

AAAS, publisher of *Science*, thanks the sponsors and supporters of the 2019 Annual Meeting.

European Commission

GREAT BRITAIN & NORTHERN IRELAND

SUBARU

Johnson & Johnson INNOVATION

Federal Ministry of Education and Research

Research in Germany Land of Ideas

Smithsonian

Topical Lectures

Materials for the Future: Self-driving Laboratories

ALÁN ASPURU-GUZIK
Professor of Chemistry and Computer Science, University of Toronto

Climate Change 2019: Finding the Accelerator Pedal

CHRISTOPHER B. FIELD
Director, Stanford Woods Institute for the Environment, Stanford University

Organs-on-Chips Technology: Drug Discovery and Development Beyond Animal Models

GERALDINE HAMILTON
President and Chief Scientific Officer, Emulate, Inc.

Secrets of Spider Webs: Unraveling the Complexity of Spider Silk Genes

CHERYL Y. HAYASHI
Director of Comparative Biology Research, American Museum of Natural History

Critical Steps Toward Modernizing Graduate STEM Education

ALAN I. LESHNER
CEO Emeritus, American Association for the Advancement of Science

Health Beyond Humanity: A Planetary Perspective On Our Past and Future

SABRINA SHOLTS
Curator of Biological Anthropology, National Museum of Natural History, Smithsonian Institution

Heredity: Our Defining Mystery

CARL ZIMMER
Columnist, *New York Times*

SARTON MEMORIAL LECTURE IN THE HISTORY AND PHILOSOPHY OF SCIENCE

Science for Grown-Ups: Assessing Past and Present Adult Informal Science Education

KAREN A. RADER
STS Director & Professor of History, Virginia Commonwealth University

JOHN P. MCGOVERN AWARD LECTURE IN THE BEHAVIORAL SCIENCES

The Gestural Origins of Language and Thought

SUSAN GOLDIN-MEADOW
Beardsley Ruml Distinguished Service Professor, University of Chicago

Updated program information will be posted on **aaas.org/meetings** as it becomes available.

Celebrating Milestones in Science!

In 1869,

Dmitri Mendeleev presented the Periodic Table, which documented the atomic weights and valence of the many elements that make up Earth, to the Russian Chemical Society.

One hundred years later, in 1969,

history was made by NASA with the first moon walk by Neil Armstrong, made possible by Apollo 11's successful landing on the moon.

On the ground in California that same year,

the first ARPANET message was sent from the University of California, Los Angeles to the Stanford Research Institute.

Scientific Sessions

Keep up with cutting edge developments in science and policy

Controlling Contagion

ANTIBIOTIC USE IN HUMANS AND ANIMALS: STRATEGIES FROM SOCIAL SCIENCE
Organized by Renata Ivanek Miojevic and Michelle Wemette, Cornell University, Ithaca, NY

EPIGENETICS: CHALLENGES FOR RESPONSIBLE RESEARCH AND INNOVATION
Organized by Vittorio Colizzi, University of Rome—Tor Vergata, Italy; Maria S. Salvato, University of Maryland School of Medicine, Baltimore, MD

INFECTIOUS DISEASE CONTROL AND PREVENTION: THE FUTURE IS INTERDISCIPLINARY
Organized by Terry O'Connor, UK Research and Innovation, Swindon, United Kingdom

INFECTIOUS DISEASES: PUSHING THE BOUNDARIES OF PHYSIOLOGY
Organized by Isabelle Boscaro-Clarke, Diamond Light Source, Didcot, United Kingdom

MATHEMATICAL MODELING OF DISEASES: TRANSLATIONAL APPROACHES
Organized by Reinhard Laubenbacher, UConn Health, Farmington, CT; Mark Alber, University of California, Riverside, CA

OPENNESS ABOUT CLINICAL TRIAL RESULTS: LESSONS FROM COMPANIES ON THE FRONT LINE
Organized by Síle Lane, Sense about Science, London, United Kingdom

THE QUEST FOR A UNIVERSAL FLU VACCINE
Organized by Karen Chandross, Sanofi, Bridgewater, NJ

THE ROLE OF STEM CELLS IN THE FUTURE OF MEDICINE
Organized by Jens Wilkinson, RIKEN, Saitama, Japan; Ayaka Nakauchi, Kyoto University, Japan

THE SILENCE OF THE FROGS
Organized by Erin Heath, AAAS, Washington, DC; Julia Smith, Association of American Universities, Washington, DC

UNDERSTANDING ANTIBIOTIC RESISTANCE: REGULATION, QUALITY, AND ACCESS
Organized by Muhammad Zaman, Boston University, MA

VIRUSES, MICROBES, AND THEIR ENTANGLED FATES
Organized by Joshua Weitz, Georgia Institute of Technology, Atlanta, GA; Adrienne Correa, Rice University, Houston, TX

Cultivating Borderless Research

BEYOND OR DESPITE POLITICAL BORDERS: SCIENCE DIPLOMACY AND THE CONSTRUCTION OF EUROPE
Organized by Pascal Griset, Sorbonne Université, Paris, France; Elke Boers, Vrije Universiteit Brussel, Brussels, Belgium

BORDER WALLS: EXCLUSIONARY AND INEFFECTIVE
Organized by Dudley Poston, Texas A&M University, College Station, TX

BUILDING GLOBAL RELATIONSHIPS THROUGH DEFENSE DIPLOMACY AND SCIENTIFIC COOPERATION
Organized by Esha Mathew, AAAS, Washington, DC; Taeyjuana Lyons, U.S. Department of Defense, Alexandria, VA

COMPETITIVE ADVANTAGE: PROTECTING THE BALANCE IN AFRICA, ASIA, AND EUROPE
Organized by Aidan Gilligan, SciCom—Making Sense of Science, Brussels, Belgium

FUNDING AGENCIES AND RESEARCH: COPING WITH GEOGRAPHIC AND DISCIPLINARY CHALLENGES
Organized by Shaun Baron, Natural Sciences and Engineering Research Council of Canada, Ottawa, Canada; Rainer Gruhlich, German Research Foundation, Washington, DC

GLOBAL SCIENCE PARTNERSHIPS AND LOCAL OUTCOMES
Organized by Elizabeth Lyons, National Science Foundation, Alexandria, VA

GLOBAL SCIENTIFIC COLLABORATIONS: NEW TRENDS
Organized by David Cheney, Technology Policy international, LLC, Silver Spring, MD; Go Ohba, National Institute of Science and Technology Policy, Tokyo, Japan

GOING GLOBAL: TOWARD A TRULY INTERNATIONAL PROCESS FOR RESEARCH INFRASTRUCTURES
Organized by Terry O'Connor, UK Research and Innovation, Swindon, United Kingdom; Peggy Pan, Chinese Academy of Sciences, Beijing, China

KILLER ROBOTS: TECHNOLOGICAL, LEGAL, AND ETHICAL CHALLENGES
Organized by Toby Walsh, University of New South Wales, Kensington, Australia

COUNTDOWN TO AFRICA'S SCIENTIFIC LIFTOFF
Organized by Neil Turok, Perimeter Institute for Theoretical Physics, Waterloo, Canada

SCIENCE DIPLOMACY: CONTRIBUTIONS TO THE RESOLUTION OF CONFLICTS AND GLOBAL INSECURITY
Organized by Karen Lips, University of Maryland, College Park, MD; Meredith Gore, Michigan State University, East Lansing, MI

SCIENCE IN AFRICA: NEW PROGRAMS AND CONTRIBUTIONS
Organized by Herman B White, Fermi National Accelerator Laboratory, Batavia, IL; Ketevi Assamagan, Brookhaven National Laboratory, Wading River, NY

SCIENCE IN THE NEW ARCTIC: THE CONVERGING OF NATURAL AND SOCIAL SCIENCES
Organized by Andrey Petrov, International Arctic Social Sciences Association, Cedar Falls, IA; Jack Kaye, NASA, Washington, DC

TREATING PEDIATRIC CANCER IN CRISIS: LESSONS FOR DELIVERING CARE
Organized by Sima Jeha, St. Jude Children's Research Hospital, Memphis, TN

Diversifying Scientific Contributions

HOW PEOPLE LEARN: A NEW LOOK
Organized by Nora Newcombe, Temple University, Philadelphia, PA; Zewelanji Serpell, Virginia Commonwealth University, Richmond, VA

ACADEMIC RESEARCH ASSESSMENT: REDUCING BIASES IN EVALUATION
Organized by Anna Hatch, American Society for Cell Biology, Bethesda, MD; Stephen Curry, Imperial College, London, United Kingdom

EFFECTIVE MENTORING: STRATEGIES AND EXPERIENCES
Organized by Maria Lund Dahlberg, National Academy of Sciences, Engineering, and Medicine, Washington, DC; Juan Gilbert, The University of Florida, Gainesville, FL

PRACTICAL REPRODUCIBLE RESEARCH: REPLICATING SCIENTIFIC DISCOVERIES
Organized by Susan Holmes, Stanford University, CA

PROMOTING GENDER EQUITY IN STEM: TOP DOWN AND BOTTOM UP APPROACHES
Organized by Christine O'Connell, Stony Brook University, NY; Merryn McKinnon, Australian National University, Canberra, Australia

REDUCING SOCIETAL POLARIZATION ABOUT SCIENCE-BASED ISSUES: PROPOSED STRATEGIES
Organized by Edward W. Maibach, George Mason University, Fairfax, VA

REFUGEES AND IMMIGRANTS: HISTORIC CONTRIBUTIONS TO STEM
Organized by Lynn Caporale, New York, NY

SCIENTIFIC LEADERSHIP PROGRAMS: PREPARING SCIENTISTS TO MAKE A DIFFERENCE
Organized by Giovanna Guerrero-Medina, Yale University & Ciencia Puerto Rico, New Haven, CT

SEXUAL MINORITIES IN THE STEM PIPELINE: FROM EDUCATION TO WORKFORCE
Organized by Rochelle Diamond, National Organization of Gay and Lesbian Scientists and Technical Professionals, Pasadena, CA

SUPPORTING UNDERGRADUATE RESEARCHERS TRANSITIONING BETWEEN 2 AND 4 YEAR COLLEGES
Organized by Linda M. Grisham, Massachusetts Bay Community College, Wellesley Hills, MA; Stacey Kiser, Lane Community College, Eugene, OR

USING ONLINE PLATFORMS TO CREATE CULTURALLY RELEVANT COMMUNITIES OF PRACTICE
Organized by Lou Woodley, AAAS, Washington, DC

WOMEN IN SCIENCE: UNDERSTANDING WHAT WORKS
Organized by Ashley Bear, National Academies of Sciences, Engineering, and Medicine, Washington, DC; Tom Rudin, Board on Higher Education and Workforce, Washington, DC

Engaging the Human Ecosystem

COMPUTATIONAL SOCIAL SCIENCE: THE NEXT DECADE
Organized by David Lazer, Northeastern University, Boston, MA

DRUG DEVELOPMENT: OVERCOMING OBSTACLES
Organized by Richard Neubig, Michigan State University, East Lansing, MI; William Beck, University of Illinois, Chicago, IL

GLOBAL NEUROSCIENCE: ACCELERATING BRAIN SCIENCE DISCOVERY
Organized by Stephanie Albin, The Kavli Foundation, Los Angeles, CA

HUMAN GENETIC RESEARCH: OVERCOMING RACE
Organized by Keegan Sawyer, National Academies of Sciences, Engineering, and Medicine, Washington, DC; Michael Yudell, Drexel University, Philadelphia, PA

MEDICINE, COMPUTER SCIENCE, AND ART: LEARNING THROUGH TECHNOLOGY
Organized by Francesca Casadio, Art Institute of Chicago, IL

MISSIONS TO MARS: UNDERSTANDING AND PREPARING TEAMS FOR THE FUTURE
Organized by Noshir Contractor, Northwestern University, Evanston, IL

NEUROBIOLOGICAL AND PUBLIC HEALTH PERSPECTIVES ON EARLY LIFE ADVERSITY
Organized by Rachel Anderson, AAAS, Washington, DC

NEURODEVELOPMENTAL DISORDERS: DEVELOPING MEDICAL TREATMENTS
Organized by Jacqueline Crawley, University of California, Davis School of Medicine, Sacramento, CA

NEUROLOGY AND PSYCHIATRY: MEETING IN THE AGING BRAIN
Organized by Irina Esterlis, National Center for Post-Traumatic Stress Disorder, West Haven, CT

NORDIC REGISTERS AND BIOBANKS: A GOLDMINE FOR PRECISION MEDICINE RESEARCH
Organized by Tor Martin Nilsen, NordForsk, Oslo, Norway

P7: A NEW PARADIGM FOR HEALTH CARE IN THE 21ST CENTURY
Organized by Ram D. Sriram, National Institute of Standards and Technology, Gaithersburg, MD; Ramesh Jain, University of California, Irvine, CA

PREVENTING SEXUAL HARASSMENT: NEXT STEPS FOR SCIENCE, ENGINEERING, AND MEDICINE
Organized by Frazier Benya, National Academies of Sciences, Engineering, and Medicine, Washington, DC

RACE, SEX, AND GENES: SHAPING BODIES, SHIFTING BOUNDARIES, CHALLENGING MYTHS
Organized by Agustín Fuentes, University of Notre Dame, IN

TEETH AS A BIOMARKER FOR ENVIRONMENTAL STRESS AND RISK OF DISEASE
Organized by Pamela Den Besten, University of California, San Francisco, CA

Expanding Tech Applications

BIG DATA: OVERCOMING CHALLENGES TO FACILITATE DATA RE-USE
Organized by Sabina Leonelli, University of Exeter, United Kingdom; Nancy J. Nersessian, Georgia Institute of Technology and Harvard University, Cambridge, MA

BLOCKCHAIN AND THE SCIENTIFIC METHOD
Organized by Kush Varshney, IBM Research, Yorktown Heights, NY; Lav Varshney, University of Illinois at Urbana-Champaign, IL

BUILDING QUANTUM COMPUTERS: WHY AND HOW
Organized by Charles Clark, Joint Quantum Institute, Gaithersburg, MD; Anice Anderson, Private Engineering Consulting, Carmel, IN

DRIVERLESS FUTURES: DISCIPLINES, SECTORS, AND GEOGRAPHIC CONSIDERATIONS
Organized by Mahmud Farooque and Jason Lloyd, Consortium for Science, Policy & Outcomes, Arizona State University, Washington, DC

EARTH'S MANTLE REVEALS EVIDENCE OF NORTH AMERICA'S STORIED GEOLOGICAL PAST
Organized by Beth Grassi and Elisabeth Nadin, EarthScope National Office, Fairbanks, AK

EARTH'S SURFACE RESPONSE TO EARTHQUAKES, VOLCANOES, AND GROUNDWATER: BULGE AND RUPTURE
Organized by Beth Grassi and Elisabeth Nadin, EarthScope National Office, Fairbanks, AK

ENABLING LONG-DURATION MANNED EXPLORATION OF SPACE: CHALLENGES AND NEW DEVELOPMENTS
Organized by Sigrid Reinsch and Sandra Dueck, NASA Ames Research Center, Moffett Field, CA

HOW NEW TECHNOLOGY AFFECTS SOCIETY
Organized by Giovanni Felici, European Research Council Executive Agency, Brussels, Belgium; Frank Kuhn, European Research Council Executive Agency, Saint-Josse-ten-Noode, Belgium

MULTI-MESSENGER ASTROPHYSICS: INSIGHTS FROM COMBINING GRAVITY AND RADIO WAVES
Organized by Eric Murphy, National Radio Astronomy Observatory, Charlottesville, VA

PUBLIC USE DATA PRODUCTS AND PRIVACY PROTECTION: EXPERIENCES FROM THE FRONT LINES
Organized by Jerome Reiter, Duke University, Durham, NC

SOLVING THE BIGGEST MYSTERIES WITH THE TINIEST PARTICLES
Organized by Kurt Riesselmann, Fermi National Accelerator Laboratory, Batavia, IL

SYNTHETIC BIOLOGY AND ADAPTIVE MANAGEMENT
Organized by Wayne Landis, Western Washington University, Bellingham, WA; Keegan Sawyer, The National Academies of Sciences, Engineering, and Medicine, Washington, DC

THE ORIGIN OF GOLD, PLATINUM, AND OTHER HEAVY ELEMENTS IN THE UNIVERSE
Organized by Reiner Kruecken and Jonathan Bagger, TRIUMF, Vancouver, Canada

THE SCIENCE OF QUANTUM INFORMATION: DISENTANGLING OPPORTUNITIES FROM THE HYPE
Organized by David Steuerman, The Kavli Foundation, Los Angeles, CA

Feeding Society

ADDRESSING HUNGER THROUGH AGRICULTURAL DISCOVERY
Organized by Elizabeth Ainsworth, USDA Agricultural Research Center, Urbana, IL

FOOD AND NUTRITION SECURITY: SCIENTIFIC PARTNERSHIPS AND OPPORTUNITIES
Organized by Volker ter Meulen, InterAcademy Partnership, Trieste, Italy; Robin Fears, European Academies' Science Advisory Council, Halle (Saale), Germany

FOOD WASTE: A SOCIETAL CONCERN
Organized by Sheril Kirshenbaum, Michigan State University, East Lansing, MI; Michael E. Webber, University of Texas at Austin, TX

GENOME EDITED LIVESTOCK AND REGULATORY STALEMATES
Organized by Alison Van Eenennaam, University of California, Davis, CA

HUMAN INTERACTIONS, BIODIVERSITY, AND SOCIO-ECOLOGICAL DYNAMICS IN DEEP TIME, PARTS I AND II
Organized by Stefani Crabtree, The Pennsylvania State University, State College, PA; Jennifer A. Dunne, Santa Fe Institute, NM

SUSTAINABLY FEEDING TEN BILLION PEOPLE
Organized by Jim Ballingall, Industry-Academia Partnership, Los Gatos, CA

Forming Connections

(IN)JUSTICE BEFORE TRIAL: UNDERSTANDING EARLY STEPS IN THE CRIMINAL JUSTICE SYSTEM
Organized by Barbara Spellman, University of Virginia, Charlottesville, VA

APPLIED SYSTEMS ANALYSIS: BRIDGING EAST AND WEST
Organized by Jan Marco Müller, International Institute for Applied Systems Analysis, Laxenburg, Austria

CENSUS 2020: RECENT CHANGES AND IMPACTS
Organized by Robin Mejia, Carnegie Mellon University, Pittsburgh, PA

FUNDING SCIENCE IN CANADA: POLITICAL INITIATIVE MEETS GRASSROOTS SUPPORT
Organized by Mehrdad Hariri, Canadian Science Policy Centre, Toronto, Canada

GOVERNMENT SCIENTISTS: ROLES AND CHALLENGES
Organized by Margaret Taylor, Lawrence Berkeley National Laboratory, Berkeley, CA

INTERNATIONAL SCIENTIFIC PARTNERSHIPS: OPPORTUNITIES AND CHALLENGES
Organized by Caroline S. Wagner, Ohio State University, Columbus, OH; Rainer Asse, American Academy of Arts and Sciences, Cambridge, MA

LEGAL PROTECTION FOR SCIENTISTS: OVERCOMING REAL AND PERCEIVED BARRIERS
Organized by Joanne P. Carney and Theresa Harris, AAAS, Washington, DC

PARTICLE PHYSICS: FRONTIER SCIENCE AND INTERNATIONAL PARTNERSHIPS
Organized by Pushpalatha Bhat, Fermi National Accelerator Laboratory, Batavia, IL

POLICYMAKERS AND COMMUNICATING SCIENCE: OPPORTUNITIES AND BEST PRACTICES
Organized by Kei Koizumi, AAAS, Washington, DC; Karen Akerlof, George Mason University, Fairfax, VA

SCIENCE DIPLOMACY AND NORTH KOREA: THE DAWN OF A NEW ERA
Organized by Mijung Jung and Soyeon Shin, National Research Council of Science & Technology, Sejong, Korea, Republic of (South)

SCIENCE INFORMING POLICY: RESPONSIBLE DECISION MAKING
Erica Goldman, National Council for Science and the Environment, Washington, DC; Stephen Posner, COMPASS, Silver Spring, MD

SCIENCE MEETS PARLIAMENT: LESSONS LEARNED
Organized by Kylie Walker and Emma Johnston, Science & Technology Australia, Canberra, Australia

Navigating Technology

ACCESS TO PRIVATE AND PROPRIETARY DATA: THEOREMS AND WORKING PROTOTYPES
Organized by Daniel Goroff, Alfred P. Sloan Foundation, New York, NY

AI AND MACHINE LEARNING: INDUSTRIAL APPLICATIONS AND SOCIAL IMPLICATIONS
Organized by Sudarsan Rachuri, U.S. Department of Energy, Washington, DC; James Garrett, Carnegie Mellon University, Pittsburgh, PA

CYBERSECURITY: TRANSCENDING PHYSICS, TECHNOLOGY, AND SOCIETY
Organized by Kevin Fu, University of Michigan, Ann Arbor, MI

HOW AI AND KNOWLEDGE CENTERS ARE CHANGING SOCIETAL VIEWS OF CRITICAL EARTH RESOURCES
Organized by Suzanne A. Pierce, Texas Advanced Computing Center, Austin, TX; Vipin Kumar, University of Minnesota, Minneapolis, MN

INEQUALITY IN NETWORKS, MINDS, AND MACHINES
Organized by Keith Payne, University of North Carolina at Chapel Hill, NC

MACHINE LEARNING AND HUMAN LANGUAGE
Organized by Cecile McKee, University of Arizona, Tucson, AZ

MACHINE LEARNING AND STATISTICS: APPLICATIONS IN GENOMICS AND COMPUTER VISION
Organized by Hal Stern, University of California, Irvine, CA

MACHINE LEARNING FOR THE AUTOMATION OF SCIENTIFIC DISCOVERY ACROSS DISCIPLINES
Organized by Ben Wender, National Academies of Sciences, Engineering, and Medicine, Washington, DC

MASS SPECTROMETRY: NEW ADVANCES IN CHEMISTRY, ARCHEOLOGY, AND PALEONTOLOGY
Organized by Timothy Cleland and Caroline Solazzo, Smithsonian Institution, Suitland, MD

MICROGRAVITY RESEARCH: TIME, SPACE, AND MATERIALS
Organized by Lynnette D. Madsen, National Science Foundation, Alexandria, VA

SOCIO-TECHNICAL CYBERSECURITY: IT'S ALL ABOUT PEOPLE
Organized by Keith Marzullo, University of Maryland, College Park, MD

Reaching Sustainable Development

ADVANCING SUSTAINABLE DEVELOPMENT GOALS THROUGH CHEMISTRY
Organized by Mary M. Kirchhoff, American Chemical Society, Washington, DC

CLIMATE IMPACTS ANALYSIS: INTEGRATING SCIENTIFIC DISCIPLINES
Organized by Juan-Carlos Ciscar, European Commission, Seville, Spain; Marton Hajdu, European Commission, Brussels, Belgium

SUSTAINABLE DEVELOPMENT GOALS: NEW SCIENCE BALANCING A HYPERCONNECTED WORLD
Organized by Sue Nichols and Jianguo (Jack) Liu, Michigan State University, East Lansing, MI

SUSTAINABLE DEVELOPMENT: IMPLEMENTING SYSTEMIC DIMENSIONS
Organized by Jan Marco Müller, International Institute for Applied Systems Analysis, Laxenburg, Austria

SUSTAINABLE INVESTING: SCIENCE, ASSET MANAGEMENT AND SUSTAINABLE DEVELOPMENT
Organized by Brad Wible, *Science* magazine, Washington, DC; Charles Vorosmarty, The City College of New York, NY

SUSTAINABLE TRANSITIONS: LEADERSHIP MODELS FOR COMMUNITY COLLABORATION
Organized by Josh Tewksbury, University of Colorado and Colorado State University, Boulder, CO

THE CARBON CYCLE: CONDITIONS AND OPPORTUNITIES FOR CAPTURE AND UTILIZATION
Organized by Carmine Marzano, European Commission, Brussels, Belgium

THE RESEARCH CLOUD: SUPPORTING THE SUSTAINABLE DEVELOPMENT GOALS
Organized by Vinny Pillay, South African Department of Science and Technology, Brussels, Belgium

TO SAVE 1,000,000 CHILDREN A YEAR: DATA, SCIENCE, AND SUSTAINABLE DEVELOPMENT
Organized by Abraham Flaxman, Institute for Health Metrics and Evaluation, Seattle, WA

Rehabilitating Habitats

EXTREME EVENT ATTRIBUTION IN THE CONTEXT OF CLIMATE CHANGE
Organized by Francis Zwiers, University of Victoria, Canada; Debbie J. Dupuis, HEC Montreal, Canada

CARBON CYCLE PROCESSES, POLICIES, AND UNCERTAINTIES: NEW PERSPECTIVES
Organized by Maureen McCarthy, University of Nevada, Reno, NV; Gyami Shrestha, U.S. Global Change Research Program, Washington, DC

CLUTCHING AT STRAWS: SCIENCE ADVICE, UNCERTAINTY, AND GLOBAL MICROPLASTIC POLLUTION
Organized by Jeremy Bray, European Commission, Brussels, Belgium

CLIMATE CHANGE: UNDERSTANDING FEEDBACK FROM NATURE, CULTURE, AND SOCIETY
Organized by Marianne Lucien, ETH Zürich, Switzerland

COMMUNITY RESPONSES TO CLIMATE CHANGE
Organized by Emily Therese Cloyd and Elana Kimbrell, AAAS, Washington, DC

ECOLOGIES OF INNOVATION: THE POTENTIAL FOR BIOTECHNOLOGY TO ADDRESS FOREST HEALTH
Organized by Kara Laney, National Academies of Sciences, Engineering, and Medicine, Washington, DC

ENVIRONMENTAL AND TEXTILE SCIENTISTS COMBATING MICROPLASTIC POLLUTION
Organized by Margaret Murphy, Washington, DC; Judith Weis, Rutgers University, Newark, NJ

ENVIRONMENTAL HEALTH DATA INTEGRATION: MEASUREMENT AND IMPACT
Organized by Andrea Hodgson and Ben Wender, The National Academies of Sciences, Engineering, and Medicine, Washington, DC

ENVIRONMENTAL MICROBIOMES: BACTERIAL AND FUNGAL COMMUNITIES IN EXTREME ECOSYSTEMS
Organized by Barbara Illman, U.S. Forest Service, Madison, WI

HOMES AT THE CENTER OF CHEMICAL EXPOSURE: UNITING CHEMISTS, ENGINEERS AND HEALTH SCIENTISTS
Organized by Glenn Morrison, University of North Carolina at Chapel Hill, NC; Jon Abbatt, University of Toronto, Canada

PLASTICS IN THE OCEANS: SOURCES, SINKS, AND SOLUTIONS
Organized by Linsey Haram and Christina Simkanin, Smithsonian Environmental Research Center, Edgewater, MD

SUSTAINABILITY SCIENCE AND PUBLIC ENGAGEMENT: INVOLVING LAND OWNERS IN RESEARCH
Organized by Kimberly La Pierre, Smithsonian Environmental Research Center, Edgewater, MD; Sally Koerner, University of North Carolina, Greensboro, NC

THE FOURTH NATIONAL CLIMATE ASSESSMENT: ADVANCING SCIENCE, INFORMING DECISIONS
Organized by Katie Reeves, U.S. Global Change Research Program, Washington, DC; Donald J. Wuebbles, University of Illinois at Urbana-Champaign, IL

TRANSBOUNDARY AIR POLLUTION: THE IMPACT OF SCIENCE ON POLICY
Organized by Terry Keating, U.S. Environmental Protection Agency, Washington, DC; Barry Lefer, NASA, Washington, DC

Sharing Science

A FEMINIST AGENDA FOR SCIENCE COMMUNICATION: NECESSARY AND TIMELY
Organized by Megan Halpern, Michigan State University, East Lansing, MI

AGRICULTURAL INNOVATION AND UPTAKE: USING A MULTI-ACTOR APPROACH
Organized by Jean-Marc Gautier, Institut de l'Elevage, Castanet-Tolosan Cedex, France; Cathy Dwyer, Scotland's Rural College, Midlothian, United Kingdom

FIGHTING FAKE NEWS: VIEWS FROM SOCIAL AND COMPUTATIONAL SCIENCE
Organized by Stephan Lewandowsky, University of Bristol, United Kingdom; Edward W. Maibach, George Mason University, Fairfax, VA

GLOBALIZATION OF THE STEM WORKFORCE: IMPLICATIONS FOR CAREERS AND INSTITUTIONS
Organized by Julia E. Melkers, Georgia Institute of Technology, Atlanta, GA; Eric Welch, Arizona State University, Phoenix, AZ

HUMAN GENETIC VARIATION AND EDUCATION: NOT A SOCIALLY NEUTRAL ENDEAVOR
Organized by Jonathan Beckwith, Harvard Medical School, Boston, MA; Kostia Bergman, Northeastern University, Boston, MA

HUMAN RESOURCE DEVELOPMENT AND DIVERSITY: INSIGHTS FROM PUBLIC AND PRIVATE SECTORS
Organized by Takashi Inutsuka, Ministry of Education, Culture, Sports, Science and Technology, Tokyo, Japan

IMPROVING ATTITUDES TOWARDS CHEMISTRY THROUGH INFORMAL SCIENCE COMMUNICATION
Organized by Larry Bell and David Sittenfeld, Museum of Science, Boston, MA

SCIENCE ENGAGEMENT WITH FAITH COMMUNITIES
Organized by Robert O'Malley, AAAS, Washington, DC

SCIENTIFIC INTEGRITY: PRINCIPLES AND BEST PRACTICES
Organized by Alison Kretser, International Life Sciences Institute North America, Washington, DC

TALKING WITHOUT SPEAKING: OVERCOMING COMMUNICATION CHALLENGES WITH TECHNOLOGY
Organized by Nan Ratner, University of Maryland, College Park, MD

TECHNOLOGY TRANSFER AND INNOVATION: RESHAPING LABORATORY-MARKET RELATIONSHIPS
Organized by Heather Evans, National Institute of Standards and Technology, Gaithersburg, MD; Anice Anderson, Private Engineering Consulting, Carmel, IN

TECHNOLOGY TRANSFER FROM BLUE SKY SCIENCE: AVOIDING PITFALLS, MAXIMIZING RETURNS
Organized by Terry O'Connor, UK Research and Innovation, Swindon, United Kingdom; Ana Godinho, CERN, Geneva, Switzerland

THE BIOLOGY OF RESILIENCE: HOW SCIENCE AND FAITH COMMUNITIES CAN WORK TOGETHER
Organized by Se Y. Kim and Curtis Baxter, AAAS, Washington, DC

YOUTUBE: FRIEND OR FOE IN COMMUNICATING ABOUT SCIENCE AND HEALTH
Organized by Erik Bucy and Asheley Landrum, Texas Tech University, Lubbock, TX

Flash Talks

Brief presentations highlighting scientific findings and programs

A Balancing Act: Navigating Jargon Use When Communicating with Citizen Scientists

The Biology of Fatherhood: Adaptive Origins, Day-to-Day Functions, and Men's Health

Developing Engagement Through Early Career Science Policy Groups

Development of a Transdisciplinary Scale to Measure Household Water Insecurity

Drivers of Data Ecologies in Genomics and the Infrastructure Sciences

Engagement and the All of Us Research Program: Meeting People Where They Are

Environmental Perceptions and Migration Decisions Within the United States Gulf Coast

Football, Disabilities, and Engineering: Customer Centered Innovation

From Pipette to Pen: A Scientist's Summer

Gut Microbiome of Bees and Wasps

Hard Lessons from the Soft Sciences on Gender in Engineering and Technology

International Partnerships and the Open Knowledge Africa Platform

Making Hands-on Biology Experiences Accessible for Everyone

March Mammal Madness: The Power of Science as Narrative

The Next NASA Golden Record Anchored in Science, Information, and Digital Literacies

Nonlinear Optical Study of Two-Dimensional Materials

Optical Markers and Biomarkers for the Risk Assessment of Oral Premalignant Lesions

Promising Genetic Research for Prognosis and Treatment of Many Cancers

Public Engagement Leadership: A Journey From Practitioner to Change Agent

#ScientistsWhoSelfie: Instagramming the Way to Public Trust

The Second Quantum Revolution

The Shape of Human Evolution

Using Policy Analysis to Engage Policymakers and the Public with Research Results

Using Science to Engage Business

Virtual Scientific Communities: Choose Development! Takes Broadening Participation to the Next Level

Career Workshops

Opportunities to gain advice and strategies from experienced STEM professionals

A Shared Vision: Overcoming the Barriers for Scientists to Communicate and Engage

Accessibility and Inclusion in STEM Education

Applying Negotiation Tools to Improve Communication and Address Conflict

Attract Recruiters and Jump-start Networking With an All Star LinkedIn Profile

Authorship Decisions: Advocating for Representation

Conducting Science Outreach Online: Social and Multimedia

Diversity + STEM = X: Solving the Equation for Higher Education and the Workforce

Exceptional Presentations in Spite of PowerPoint: How to Communicate Science in the Digital Age

Formulate to Communicate: Heroes, Improv, and Science Storytelling

From a Moment to a Movement: Building Infrastructure to Sustain Scientist Advocacy

Influencing Policy: Opportunities for Scientists and Engineers

LGBTQ+ in Academia and the Workplace: Rights and the Law

Making the Most of the First Two Years On the Tenure Track

Overcoming the STEM-Policy Divide with Fellowships in State Governments

Research Funding: Exploring Programs in the European Union

Scholarly Publishing: Avoiding Pitfalls and Showing Impact

Scientists Engaged in Human Rights: Professional Societies Offer Opportunities

Scientists Who Draw Comics: The Double Life of Visual Science Communicators

The Gender Dimension of Science Advice

The Global Entrepreneur: How to Harness International Innovation

YouTube as Outreach: How to Document Research using Social Media and Video

Sponsorship and Exhibit Spaces Still Available

Enhance your visibility at the AAAS Annual Meeting. Don't miss this chance for your research, products, and services to grab the attention of researchers, scientists, engineers, journalists, policymakers, educators, students, hiring managers, and funding agencies.

Visit **aaas.org/meetings** for more information.

Advance registration rates are available now through **January 22, 2019.**
On-site registration rates will apply thereafter.

	Advance Rates for AAAS Member for members in good standing	Advance Rates for Non-Member for all other attendees	On-site Rates after 1/22/2019 AAAS Member/Non-Member
General Attendee	$310	$440	$380/480
Postdoc	$135	$360	$135/380
K-12 Teacher	$135	$360	$135/380
Retired Professional	$250	$360	$295/380
Student	$65	$95	$75/105
One-Day	$175	$220	$200/240

Check **aaas.org/meetings** for the most up-to-date information.

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

WAYNE STATE UNIVERSITY

IBio

Faculty Positions in Translational Neurosciences: Addiction

As part of a broad institutional initiative in Integrative Biosciences, the Translational Neuroscience Initiative (TNI) at Wayne State University (WSU) is recruiting up to three faculty positions (tenured or tenure-track, open rank). The program in translational neuroscience fosters interdisciplinary, integrative, and collaborative approaches to understand the basic neurobiology of addiction and its clinical application in substance use (addictive) disorders. Primary focus areas are as follows.

- Pharmacological, neuromodulatory, or behavioral interventions that target novel mechanisms for treating substance use disorders
- Genetic, biomarker, and biomedical imaging approaches centered on addiction-related processes or behaviors
- Leveraging this knowledge to direct precision approaches that can improve outcomes in diverse substance-using populations (e.g. those with psychiatric, pain, or other co-occurring conditions)

The TNI serves to promote integrative and collaborative research in basic and clinical neuroscience and is situated in the new 200,000 sq. ft. Integrative Biosciences Center (IBio) that houses coordinated inter- and trans-disciplinary research teams, and a Clinical Research Center. IBio is a fulcrum for leading-edge technology platforms and specialized resources in support of advanced studies in precision medicine. Through research, clinical care, community engagement, and education, the TNI team of researchers and community partners seek to discover, investigate, and solve complex health problems that affect the human nervous system. Successful faculty candidates will have a Ph.D., M.D., or equivalent degree in biomedical sciences relevant to neurosciences with evidence of peer recognition in the field, a commitment to excellence in research education and training, and the ability to engage with broader neuroscience themes for the purpose of achieving transformative and translational research gains. Applicants are expected to have already established extramural research funding or to be on a clear path to secure extramural funding in support of their research programs. Faculty recruits will integrate with departments and colleges or schools consistent with their areas of expertise and shared interests and engage in all aspects of our academic mission, including research, training, instruction and service.

Competitive recruitment packages are available with salary and faculty rank based on qualifications. Applicants are encouraged to apply to **posting 043730** and **043718** through the WSU Online Hiring System **https://jobs.wayne.edu**. Applications will be accepted until positions are filled, but for full consideration this fall, application materials should be submitted by **November 30, 2018**. Applications should include a *curriculum vitae* and a brief narrative cover letter addressed to Mark Greenwald, Ph.D. and the Vice President for Research indicating the applicant's potential for research synergy within the TNI and the broader institutional initiative in integrative biosciences.

Wayne State University is a premier, public, urban research university located in the heart of Detroit where students from all backgrounds are offered a rich, high quality education. Our deep-rooted commitment to excellence, collaboration, integrity, diversity and inclusion creates exceptional educational opportunities preparing students for success in a diverse, global society. WSU encourages applications from women, people of color and other underrepresented people. WSU is an affirmative action/equal opportunity employer.

Founded in 1868, Wayne State University offers a range of academic programs through 13 schools and colleges to nearly 28,000 students. The campus in Midtown Detroit comprises 100 buildings over 200 acres including the School of Medicine, the Eugene Applebaum College of Pharmacy and Health Sciences and the College of Nursing. The university is home to the Perinatology Research Branch of the National Institutes of Health, the Karmanos Cancer Center, a National Cancer Institute-designated comprehensive cancer center, and a National Institute of Environmental Health Sciences Core Center - *Center for Urban Responses to Environmental Stressors (CURES)*.

Top employers embrace change based on a stable foundation

Results are in from the 16th annual Science Careers Top Employers Survey. Responses from employees in the biotechnology and pharmaceutical industry reflect growing interest in new R&D data sources and technologies, general uncertainty about drug pricing and regulations, and a tumultuous year in global politics. Companies chosen as top employers are those who proactively make necessary changes while holding fast to their fundamental principles. **By Chris Tachibana**

Like the biotech and pharma industry itself, the annual *Science* Careers Top Employers Survey continues to change and grow. This year, more than 8,000 responded, the most in the history of the survey, up from 6,950 last year. Of more than 180 companies mentioned frequently by survey participants, 20 emerged as top employers.

Survey respondents were mainly from North America (63%), followed by Europe (24%) and the Asia/Pacific Rim (9%). Most were industry employees; 93% work in a biotech, biopharma, or pharmaceutical company. Although 80% were age 30 or older, and two-thirds reported 10 or more years of work experience, 76% said they had not yet reached their career peak.

Innovation was the leading driver for top employers. Work culture, respect for employees, and social responsibility were also highly valued. An unusual feature in this year's survey compared to recent years was the emergence of "top leadership that successfully makes changes needed to keep the organization moving in the right direction" as a characteristic of top employers.

Comments from respondents reflected a year of prominent political news, impactful elections worldwide, and heated public discussions about health care reform, drug pricing, and industry regulations. The announcement by Amazon, JPMorgan Chase, and Berkshire Hathaway of a joint venture to create a technology-based health care company may have been on survey respondents' minds when they noted new data sources, analytic methods, and automation as industry changes. Other comments covered market trends, including mergers and acquisitions and the rise of biosimilars and generic drugs, and new R&D avenues such as gene editing.

Against that backdrop, representatives from top companies explain how their organizations respond to an ever-changing industry while holding to "true north." A common theme was making rational, data-driven decisions and taking actions that reflect their company's location, employee base, and foundational values.

Stability at the top

For the third straight year and the sixth time since 2011, **Regeneron Pharmaceuticals** in Tarrytown, New York, was chosen as the No. 1 employer. Chief Scientific Officer and President George Yancopoulos is enthusiastic about the top ranking, saying, "It never gets old." He attributes the continuing recognition to two major factors: "We're the only major biopharma company started by and still run by scientists after 30 years, and we continue to build on our innovation, particularly in genetics."

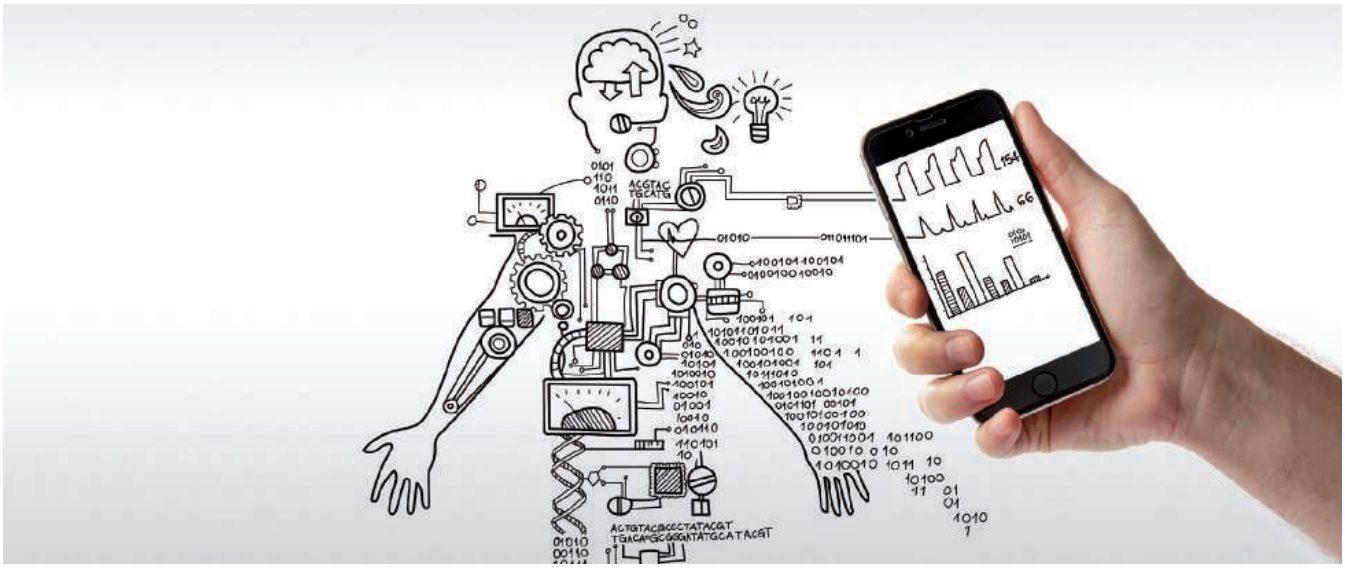
Citing the company principle of "doing well by doing good," Yancopoulos adds, "We believe that if we do the right thing based on our science, then we'll do fine from the business standpoint." He names Dupixent, an atopic dermatitis drug from Regeneron and developmental partner Sanofi, based in Paris, France, as an example. It wasn't initially predicted to be profitable, but is becoming a success, and its clinical trial data shows promise for asthma and other allergic diseases as well.

Moreover, Regeneron's leadership is not averse to change. "One of our strengths," Yancopoulos says, "is that we're willing to adapt or die. That's how we survived for 20 years before our first product approval." The company previously focused on in-house discoveries, many of which were later developed or commercialized with other companies, but is now increasingly engaged in external partnerships.

Brian Zambrowicz, Regeneron's vice president of functional genomics and chief of its *VelociGene* Operations, says the company was one of the first to work with transgenic mice and continues to build on that technology. Recent expansions in other areas include CRISPR-based gene-editing therapies through a partnership with Intellia; RNA-interference therapeutics with Alnylam; and treatments for hearing loss with Decibel Therapeutics.

Company leaders expect future discoveries from the Regeneron Genetics Center (RGC), a company subsidiary focused on genotype–phenotype data, which Yancopoulos cont.>

Upcoming features
Artificial Intelligence–November 30 ■ Postdocs 1: Careers in Diversity and Inclusion–February 1 ■ Cancer Research–March 15



Make an impact.
Code4*life*

Have you ever wondered what we could achieve if we bring together the best expertise in molecular biology, modern diagnostics and smart analytics?
How might patients' lives be changed for the better?

Let yourself be inspired by Lisa's story:

Or go directly to: go.roche.com/lisas_story

In addition to our existing strengths in oncology, immunology, infectious diseases, ophthalmology and neuroscience, we are investing into cutting-edge applications of artificial intelligence and data science to make truly Personalised Healthcare a reality.

Explore your future career in Personalised Healthcare at Roche.

The next step is yours.
code4life.roche.com

Roche

TOP TWENTY EMPLOYERS

2018 Rank	2017 Rank	Employer (global headquarters)	Innovative leader in the industry	Work culture values aligned	Treats employees with respect	Is socially responsible	Makes changes needed
1	1	Regeneron (Tarrytown, NY)	✔			✔	✔
2	–	Incyte (Wilmington, DE)	✔	✔	✔		
3	2	Novozymes (Copenhagen, Denmark)		✔	✔	✔	
4	6	Moderna (Cambridge, MA)	✔			✔	✔
5	4	Merck KGaA (Darmstadt, Germany)	✔	✔		✔	
6	3	Vertex Pharmaceuticals (Boston, MA)	✔	✔			✔
7	9	Biocon (Bengaluru, India)	✔			✔	✔
8	5	Novo Nordisk (Bagsværd, Denmark)		✔	✔	✔	
9	7	Genentech (South San Francisco, CA)	✔	✔	✔		
10	10	AbbVie (North Chicago, IL)	✔	✔		✔	
11	11	AstraZeneca/MedImmune (Cambridge, UK)		✔	✔	✔	
12	13	Roche – excluding Genentech (Basel, Switzerland)	✔	✔	✔		
13	14	Novartis (Basel, Switzerland)	✔		✔	✔	
14	12	Syngenta (Basel, Switzerland)		✔	✔	✔	
15	16	Boehringer Ingelheim (Ingelheim, Germany)		✔	✔	✔	
16	8	Eli Lilly and Company (Indianapolis, IN)		✔	✔	✔	
17	19	Celgene Corporation (Summit, NJ)	✔	✔	✔		
18	15	Abbott (Abbott Park, IL)			✔	✔	✔
19	–	Bayer (Leverkusen, Germany)			✔	✔	✔
20	17	Merck & Company (Kenilworth, NJ)	✔			✔	✔

The 20 companies with the best reputations as employers and the top three driving characteristics for each company, according to respondents in the 2018 survey undertaken for the *Science*/AAAS Custom Publishing Office. The companies without a 2017 rank did not receive enough mentions to qualify or did not receive a high enough ranking during the 2017 survey.

describes as "taking human genetics to the space-age level." RGC partners with health organizations that use clinical data or biobanks with tissue samples.

RGC's activities align with survey respondents' comments about automation and industry trends in artificial intelligence and machine learning. These methods require big data in order to train algorithms to detect patterns, Yancopoulos says, and generating those large datasets is an RGC goal. To expand RGC's database of exome sequences linked to over 300,000 individual health records, the center is leading a consortium of pharmaceutical companies, including AbbVie, AstraZeneca, Biogen, and Pfizer, to perform exome sequencing on half a million samples from UK Biobank by 2019. Each company provides USD 10 million to sequence samples linked to clinical and imaging data from study participants. Consortium members will have access to the data six to 12 months before public release.

At the No. 2 spot is a newcomer to the top employer rankings—Delaware-based pharmaceutical company **Incyte,** which also earned high marks for innovation. To CEO Hervé Hoppenot, the company's main innovation is its model: "We're a large research center with a small company on top of it. We do discovery and basic science and have a track record of developing our own medicines." Incyte's drugs focus on unmet patient needs in cancer and other conditions such as rheumatoid arthritis. The company is unique, Hoppenot says, because of its large, diverse portfolio of projects addressing different diseases, stages, mechanisms, targets, and therapeutic areas. "We have multiple ways to get to where we want to go. We don't rely on one idea about what Incyte will be in the future."

Incyte's leaders say its business decisions are rooted in scientific results. "The foundation of Incyte is medicinal chemistry and biology," says Group Vice President for U.S. Medical Affairs Peg Squier. Adds Hoppenot, "We love our chemists and biologists. They make the difference at our company. At our town halls, we have chemists give presentations, and while it's sometimes outside of some people's expertise, we think it's important that everybody knows how our projects begin."

Another unique aspect of Incyte is its collaborations, says Squier. For example, it offers drugs for investigator-initiated clinical trials earlier than other companies—right after a molecule has undergone preliminary safety profiling rather than at first regulatory cont.>

Science & Diplomacy

Science & Diplomacy provides an open access forum for rigorous thought, analysis, and insight to serve stakeholders who develop, implement, or teach all aspects of science and diplomacy. Learn more about the latest ideas in science diplomacy and receive regular updates by following @SciDip on Twitter, liking the quarterly's page on Facebook (www.facebook.com/sciencediplomacy), and registering for free at **www.sciencediplomacy.org/user/register**.

WWW.SCIENCEDIPLOMACY.ORG

Science & Diplomacy is published by the Center for Science Diplomacy of the American Association for the Advancement of Science (AAAS), the world's largest general scientific society.

SCIENCE & DIPLOMACY

AAAS
ADVANCING SCIENCE SERVING SOCIETY

DEMOGRAPHICS

GENDER:
50% Male, 45% Female, 5% No response

EXPERIENCE:
67% have 10 or more years work experience

HIGHEST DEGREE EARNED:
28% Doctorate, 31% Master's, 33% Bachelor's, 8% Other

COMPANY TYPE:
24% Pharma, 27% Biotech, 42% Biopharma, 2% University, 5% Other; More than 9 out of 10 work in private industry

NATURE OF WORK:
24% Development, 16% Applied Research, 11% Basic Research, 9% Administration/Executive, 12% QA/QC/Regulatory Affairs, 10% Production, 18% Other (respondents were able to choose more than one response)

GEOGRAPHY:
63% from North America, 24% from Europe, 9% from Asia/Pacific, 4% from rest of world

approval. Squier also cites single-trial collaborations with other companies that test the effectiveness of combined treatments. "The biology is telling us we need to look at combining therapies," she says, "so we open partnerships with other companies, sometimes just for one clinical trial to test combining our molecules."

Incyte is followed in the survey by No. 3 **Novozymes,** a global biotechnology company headquartered in Bagsværd outside of Copenhagen, Denmark. Moving from sixth place in 2017 to fourth in 2018 is Massachusetts-based **Moderna Therapeutics,** which also scored high in innovation. Moderna is developing a new class of medicines using messenger RNAs (mRNAs) instead of proteins, peptides, or small molecules—so the anticipated products themselves are the innovation. "Our science is inherently novel," says President Stephen Hoge. "There's no roadmap for how to build an mRNA company. We have to invent as we go."

Innovation is in the eye of the beholder and doesn't necessarily equate with drug discovery. A notable contrast to what other companies feel makes them innovative is found at **Biocon** in Bangalore, India, which was No. 7 after **Merck KGaA** and **Vertex Pharmaceuticals.** Biocon is known for success in generic drugs and biosimilars.

"Innovation means many things to many people," says Head of R&D Narendra Chirmule. "For mature companies, it means new targets and new therapies. But it can also mean finding new ways to do drug development and be efficient." Biocon focuses on process innovation, he says, and is moving into molecule discovery.

Process innovation at Biocon stems from its employees' broad knowledge. "If a process requires cells to increase yield," Chirmule says, "then our chemical engineers know enough cell biology to make changes through biological processes."

Culture: Context matters

Respect for employees and a work culture aligned with employee values are consistently endorsed characteristics in the top employer surveys. Comparing companies illustrates how these features are intertwined and how company culture influences the workforce and vice versa.

Chirmule describes Biocon as having "a culture of youth," with many R&D employees under 35. In addition, he says, "India's culture is about diversity." In that environment, Biocon encourages individual expression. At an annual company talent show, 80% of employees participate. "It invigorates the company and emphasizes our diversity," Chirmule says. "It shows that people have ability in both science and arts, and have diverse experiences, and it encourages a culture of innovation."

At Biocon, work–life balance means "the intermixing of science and creative arts and work and life," Chirmule says. "That's what a human being is, so we're creating a culture that lets people express all aspects of themselves." He adds, "In this day and age it's not about working from 9 to 5 and going home. Work–life balance mixes throughout the day." To stay inspired and add some fun to the workday, Chirmule keeps band equipment, including a drum set, in his office.

At Incyte, respect for individuals and work–life balance is tailored to a different employee population. Executive Vice President of Human Resources Paula Swain is not surprised that the 2018 survey found only 17% of respondents planning to seek a new job in the next year. Incyte tends to hire mid-career employees and has a stable workforce, she says. Of about 60 people who started the company in 2002, including Swain, more than 50 are still there.

Like Biocon, Incyte strives for a work culture that treats employees as individuals. At Incyte, work–life balance translates into good pay, benefits, and flexibility in accommodating individual needs. The company covers 100% of medical insurance for employees. Based on employee input, Incyte initiated paid maternity and paternity leave, including for adoption. It subsidizes a concierge service for employees for tasks such as running errands, planning vacations, and walking dogs.

"We have guidelines instead of being driven by strict rules, regulations, and policies that restrict how we handle situations," Swain says. "We focus on people as individuals." An example is offering part-time employment to people approaching retirement who aren't ready to stop working altogether. One way that Swain and Hoppenot keep in touch with employee needs is having lunch with all new employees six months after they are hired, to hear what they like and don't like about their jobs. "We're responsive to employee needs about what would make their work more pleasant, but sometimes we say no," Swain says. "We buck the trend on working from home. We want people to have flexibility but like having them interact."

Moderna's 2018 survey findings indicate that innovation unquestionably drove its employer ranking. This emphasis shapes the corporate culture. Hoge names four core company values as "bold, curious, relentless, and collaborative," and sees some combination of these traits in every employee. Being located in Cambridge, Massachusetts, also shapes Moderna's work culture. "We're pioneering something," Hoge says. "There's a bit of self-selection for people who are drawn to that kind of challenge, for which there's no guarantee of success—people who want to do things that have never been done before."

Social responsibility: An expectation

A consistent message from the survey is that employees value and expect social responsibility at their workplace. This trait has driven selection as a top employer for more than 10 years. As with work culture and innovation, corporate responsibility programs reflect a company's location, employee characteristics, and ethos.

DRIVING CHARACTERISTICS OF TOP EMPLOYERS

2018	2017
1. Innovative leader in the industry	1. Innovative leader in the industry
2. Work culture values aligned	2. Work culture values aligned
3. Treats employees with respect	3. Treats employees with respect
4. Is socially responsible	4. Is socially responsible
5. Makes changes needed	5. Has loyal employees
	6. Has clear vision

Driving characteristics are listed in descending order of impact on overall employer rankings. Colored text indicates the characteristics in common for the two years.

Social responsibility was one factor that made Regeneron the most highly regarded employer in 2018. Vice President of Corporate Communications and Citizenship Hala Mirza says corporate responsibility needs to be an authentic extension of how a company operates. "Companies can't run the commercial side of their business unethically and tack on social responsibility as a mitigating factor," she says. In other words, "You can't behave one way in business and another in corporate citizenship."

Regeneron integrates citizenship with its business and science, for example, applying its *VelociSuite* technologies for generating human antibodies to develop medicines for infectious diseases, including Zika and Middle East Respiratory Syndrome, with support from U.S. government agencies. On the business side, Regeneron's Ebola drug, developed using *VelociSuite*, earned orphan drug designation from the U.S. Food and Drug Administration.

The Regeneron priorities for corporate citizenship are supporting STEM (science, technology, engineering, and mathematics) education by developing top scientific talent, improving scientific teaching, and building awareness of scientific careers at all levels, Mirza says. "Our overall goal is elevating science in society. We envision a world in which scientists are heroes. We believe if we support them, they will solve the great problems of our time."

A keystone project is the national Regeneron Science Talent Search, which honors 300 top high school students every year, supported by an annual USD 10 million contribution from the company. Nationally and regionally, Regeneron promotes science education, especially in underprivileged communities, with mentorship, teacher training, and programs for hands-on science exploration. In 2017, Regeneron launched a company-wide Day for Doing Good, on which more than 50% of employees used their workday to participate in corporate service projects "from STEM programs to rebuilding bridges," says Mirza. Part of the employee benefit package, she adds, is paid leave for volunteering.

In addition to a paid volunteer day every year, Incyte gives employees a matching gift account, Swain says. The company matches employee donations to a charity dollar-for-dollar, up to USD 1,000 a year. Corporate giving through the Incyte Charitable Giving Foundation supports organizations in Delaware, Incyte's home state.

Biocon's corporate responsibility policies also stem from the company's vision and location, says Head of Human Resources Amitava Saha. "The cost of long-term therapy is incredibly high in most countries," he says, "particularly in a third-world economy where government support is limited. It's almost prohibitive for some patients. If you can make it affordable, it can save a lot of lives."

The Biocon Foundation helps people in remote areas by providing free or low-cost medical care, primary education, and community-infrastructure building. Employees are encouraged to take part in foundation initiatives and can use a workday for community service.

Biocon social responsibility efforts focus on education to build the talent pool in India for the biotech industry, Saha says. The company trains university graduates at the Biocon Academy, which he calls "a finishing school for biotech grads." They get classroom training and experiential learning in an industry setting. More than 300 students have been through the program, with fees subsidized up to 75% by Biocon. The academy has a 100% placement record, with graduates employed at Biocon and across the industry, Saha says. To reach a larger population of students, the academy also has refresher courses for instructors at biotech colleges.

Social responsibility boosted **Boehringer Ingelheim** to No. 15, above other organizations with similar survey rankings. Karen Iannella, president of the BI Cares Foundation, the company's philanthropic arm, says the foundation's programs focus on underserved communities. Two programs provide company medicines to people in need. A patient-assistance program gives Boehringer Ingelheim medicines for free to uninsured and underinsured U.S. patients. A product donation program does the same internationally through nonprofit partners, as part of disaster relief, for example. The foundation also contributes to community STEM education programs.

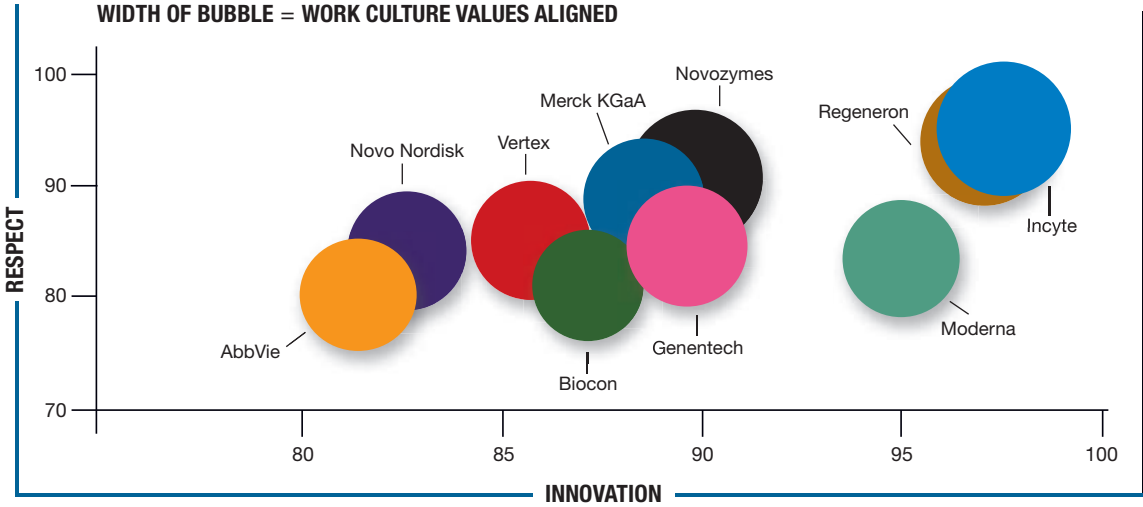
Change: A constant

The 2018 survey results represent an important cultural moment. One of the top five criteria for employers was "has top leadership that successfully makes changes." This has not been a top priority in recent years. Chirmule understands why this feature might be valued by Biocon's employees. "I interact with a lot of younger and junior staff whose role models are creators of disruptive technologies like Facebook, Tesla, or Uber," he says. "Maybe an element of the younger staff is asking, 'What is our company doing to become that innovative?'"

At Moderna, Chief Digital Officer Marcello Damiani heads a team that is digitizing all company operations. He started from scratch about three years ago with an executive-level mandate to make digitization "part and parcel of our strategy," he says. His team developed a cloud-based platform that integrates purchased software with in-house programs and links all departments, from finances to legal to R&D.

"Data entered in one system flows to all others," Damiani says, which creates efficiency. Users don't need to manage their data in individual spreadsheets. Fewer people are needed to oversee a centralized database compared to a siloed information system, he says, and researchers can easily access data across projects or clinical trials. His team has responsibility for all informatics, automation, and traditional information technology services.

The company's digitized operations include the Moderna Drug Design Studio, which supports company scientists in designing and ordering mRNA for their research. The studio suggests the best mRNA to make desired proteins, and a fully cont.>



Comparison of the top 10 companies on the basis of the top three drivers (scored out of 100):

- **Innovative leader in the industry** (*x*-axis),
- **Treats employees with respect** (*y*-axis),
- **Work culture values aligned** (bubble width).

automated central lab generates the mRNA, Damiani explains. "We collect data at each step," he says, "to learn the best way to design mRNA so the next iteration learns from the previous one." His team is now working at the company's recently opened manufacturing facility on developing digital and technical innovations that will help advance clinical programs.

One reason that survey respondents commented on mergers and acquisitions, biosimilars and generics, and big data, may be high-profile recent news—much of it involving **Roche,** this year at No. 12. Patents on three of the company's highest-earning drugs expire in the next few years, and earlier this year the company acquired Flatiron Health and Foundation Medicine.

Flatiron Health's comprehensive cancer-care software services, including electronic health records, collect patient data from hundreds of U.S. cancer-care practices. Foundation Medicine focuses on cancer gene profiling. A partnership between the two companies created the Clinico-Genomic Database, holding anonymized longitudinal clinical data on more than 20,000 cancer patients linked to their genomic profiles.

William Pao, Roche's Head of Pharma Research and Early Development and a member of its executive committee, says the company will use Flatiron Health and Foundation Medicine data in multiple ways. Applications include analyzing treatments that patients are currently getting and the resulting clinical outcomes. Roche scientists are testing how their R&D analyses compare to results using real-world clinical data from patients getting standard care.

Researchers in Pao's unit are developing digital biomarkers via a smartphone app, for patients with neuromuscular disorders such as Parkinson's disease. The app tracks symptoms over time, including development of tremors and changes in gait and voice, to gather quantitative data on how patients respond to experimental treatments, says Pao.

The focus on digital development is part of continuous learning by Roche, notes Pao. The company is 120 years old, still family owned, and has a reputation for traditional drug-and-diagnostics development. Although Roche strives for a friendly work environment and has low employee turnover, says Pao, he emphasizes that "change is our only constant."

In these turbulent times, employees seem to value companies that address uncertainties head-on. Regeneron's answer to rising drug prices was a deal with a pharmacy benefit manager to reduce the prices of its cholesterol-lowering drug in exchange for simplified patient access and physician reimbursement processes. As long as high-level decisions are consistent with a company's stated mission and its work culture, employees will understand, says Zambrowicz. Regeneron founders Yancopoulos and Len Schleifer "set the scientific focus and ethical standards and are outspoken about doing the right thing and leading by example. They set the course long ago and maintain it, so people have confidence in their leadership," he says.

Communicating decisions quickly and effectively is arguably as important to employees as making needed changes. Top employers will try to present the data and reasoning behind tough executive decisions. Especially in a science-driven industry, employees can appreciate that approach. "Focusing on science and data eliminates a lot of politics," Zambrowicz says. "If decisions are made in a rational way and driven by science, it's clear why they're made. People are comfortable with decisions based on data."

Incyte was highly rated as an employer with a work culture that respects employees, despite a setback with a melanoma drug that happened while the survey was in the field. Hoppenot attributes this outcome to the company's long-term view. "The sequence of success and failure is inherently part of the scientific process," he says. "We create a culture that accepts a level of uncertainty and organize our activities knowing that some of our projects will not work as we hoped, but that others will succeed." The strategy of running multiple, varied projects simultaneously is a way to increase the chances of success, adds Hoppenot.

Squier notes that as Incyte monitors their data, its leaders are thinking of how to respond. They aim to keep organizational bureaucracy low, to facilitate efficient decision-making. "We have contingency plans," Squier says, "so when we get a result, we don't slow down, but can change lanes if we need to. People feel that we are ready to handle whatever results come." The goal, she says, is for leaders to be straightforward with employees, so they feel like they are part of whatever decisions are made.

Yancopoulos acknowledges the uniqueness of this particular time in history: "I understand why people are concerned," he says, "We live in an uncertain time regarding political, business, and other points of view. Our answer is to continue to do the right thing, operating our business as ethically as possible and knowing the system should appropriately reward innovative companies like us."

Chris Tachibana is a science writer based in Seattle, USA, and Copenhagen, Denmark.

ETH zürich

Professors or Assistant Professors (Tenure Track) of Mechanical and Process Engineering

→ The Department of Mechanical and Process Engineering (www.mavt.ethz.ch) at ETH Zurich invites applications for the above-mentioned positions.

→ Successful candidates must have exceptional accomplishments and future potential in an area of mechanical and process engineering. The new professors should demonstrate an excellent international record of research achievements in engineering and/or natural sciences. They should have a strong motivation and indisputable commitment to undergraduate (in German or English) and graduate education (in English). The Department of Mechanical and Process Engineering is committed to promoting interdisciplinary and cutting-edge research - covering the full range from fundamental to applied - and working to meet societal challenges. Successful candidates should hold a PhD or equivalent degree in mechanical engineering, process/chemical engineering or a related field at the beginning of employment.

→ Assistant professorships have been established to promote the careers of younger scientists. ETH Zurich implements a tenure track system equivalent to other top international universities. The level of the appointment will depend on the successful candidate's qualifications.

→ **Please apply online: www.facultyaffairs.ethz.ch**

→ Applications should include a curriculum vitae, a list of publications and projects, a statement of future research and teaching interests and a description of the three most important achievements. The letter of application should be addressed **to the President of ETH Zurich, Prof. Dr. Lino Guzzella. Submissions will be reviewed starting on 1 December 2018, but applications are welcome until the positions are filled.** ETH Zurich is an equal opportunity and family friendly employer and is responsive to the needs of dual career couples. We specifically encourage women to apply.

Assistant Professor of Wildlife Habitat Ecology

The University of California at Davis is pleased to announce the recruitment for a tenure track faculty position in wildlife habitat ecology. Candidates must have the ability to develop a rigorous, extramurally funded research program that focuses on habitat ecology of vertebrate species at local, landscape, or regional scales. Research should focus on spatio temporal patterns in wildlife distributions, anthropogenic drivers of habitat use, and/or habitat management strategies, using state of the art methods and tools. The successful candidate will join the Department of Wildlife, Fish, and Conservation Biology in the College of Agricultural and Environmental Sciences at the rank of Assistant Professor. Criteria for appointment include: a Ph.D. or equivalent degree in a biological discipline relevant to wildlife habitat ecology, a record of excellence in scholarly research, and demonstrable potential to establish a competitively funded research program. The appointee will be responsible for teaching undergraduate courses in habitat and spatial ecology, as well as other departmental courses, including field courses. The appointee will also be actively involved in undergraduate advising, curricular development, and department and university service. The appointee is also expected to guide and mentor graduate students and participate in research and outreach/engagement programs consistent with the mission of the CA Agricultural Experiment Station.

Applicants should submit materials via the following website:

https://recruit.ucdavis.edu. Additional inquiries can be directed to Associate Cooperative Extension Specialist Roger Baldwin, Recruitment Committee Chair, Department of Wildlife, Fish, and Conservation Biology, One Shields Ave., University of California, Davis, CA 95616, Tel (530) 752-4551, FAX (530) 752-4154, email: mailto:rabaldwin@ucdavis.edu. The position will remain open until filled but to ensure consideration, applications should be received by December 1, 2018.

UC Davis is an affirmative action/equal employment opportunity employer and is dedicated to recruiting a diverse faculty community. We welcome all qualified applicants to apply, including women, minorities, veterans, and individuals with disabilities.

To apply, visit **http://apptrkr.com/1306256**

WAYNE STATE UNIVERSITY

Tenure/Tenure Track Faculty Position in Immunology

The Department of Biological Sciences at Wayne State University invites applications for a tenure-track opening for a researcher studying Immunology, broadly considered. Rank will be dependent upon qualifications. Preference will be given to candidates doing innovative research using model organisms in areas complementing existing departmental strengths in systems biology, development, evolution, chromatin biology, gene expression and host/pathogen interactions. Applicants must have a Ph.D. degree, postdoctoral experience and an outstanding record of research achievement. Successful applicants are expected to establish and maintain a vigorous, externally funded research program, participate in graduate and undergraduate education and perform service at the departmental, college and university level. We offer state-of-the-art research facilities and highly competitive start-up packages.

Wayne State University is a premier, public, urban research university located in the heart of Detroit where students from all backgrounds are offered a rich, high quality education. Our deep-rooted commitment to excellence, collaboration, integrity, diversity and inclusion creates exceptional educational opportunities preparing students for success in a diverse, global society. WSU encourages applications from women, people of color and other underrepresented people. WSU is an Affirmative Action/Equal Opportunity Employer.

Please submit a cover letter, curriculum vitae and 2-page statement of research plans on-line at jobs.wayne.edu (posting # 043892) by **November 16, 2018** for full consideration. Three letters of reference should be sent to: **Faculty Search Committee, Department of Biological Sciences, Wayne State University, Detroit MI 48202 (av3459@wayne.edu)**. Applications will be considered only when all materials have been received.

IUSSTF
Indo-U.S. Science & Technology Forum
New Delhi, India

Vacancy Announcement:
POSITION OF EXECUTIVE DIRECTOR

The governments of the United States of America and India, through an inter-governmental agreement, jointly established the Indo-U.S. Science & Technology Forum (IUSSTF) **http://iusstf.org** in the year 2000 as an autonomous bilateral organization located in New Delhi, India. Its mission is to promote science & technology (S&T), innovation and entrepreneurship collaborations between the United States and India.

IUSSTF invites applications for the position of Executive Director as per the details given below:

The applicant should be an Indian or U.S. citizen only. Dual citizens are also acceptable.

Essential Qualification: Ph.D or Masters in the Sciences, Engineering and Medicine.

Desirable Qualification: Degree in Management.

Tenure of appointment: 3 years with possibility of extension(s).

Age Limit: The Age limit should be up to the age of 56 years at the time of notification of the position. The cut-off date for determining the age limit shall be **November 30th, 2018**.

Compensation Package: Salary will be commensurate with experience and corresponding to the position of Scientist "G" or "H" prevailing in Scientific Ministries / Departments in India.

The selected candidate will be offered appropriate position corresponding to either Scientist "G" or "H" depending upon qualification, experience and other credentials.

Application procedure: Interested persons should send their applications to **recruitments@indousstf.org** in the prescribed IUSSTF application format with a 500 words vision statement on the future of the IUSSTF and at least 3 references.

Application deadline is 23:59 hrs. IST, November 30th, 2018.

For details on the position description and prescribed application format Please visit the **"Announcements"** section at **http://iusstf.org**

THE UNIVERSITY OF UTAH

The new **Cluster in Evolutionary Genetics and Genomics at the University of Utah** (http://www.ucegg.org), a joint initiative between the Human Genetics and Biology Departments, invites applications for a tenure-track position based within the Department of Human Genetics (School of Medicine). Individuals hired under this initiative will complement related hires in the School of Biological Sciences (College of Science). Successful applicants will join a vibrant intellectual community across the University of Utah working at the interface of genomics, computational biology, and molecular biology. The Cluster will build on substantial, ongoing investments in genome science and experimental basic science at the University of Utah, a unique and inclusive community at the foothills of the Wasatch Mountains with nearby world-class cultural and recreational opportunities.

Competitive candidates will demonstrate an ability to develop independent and vigorous research programs. Applications by investigators, both experimental and theoretical, in all areas of evolutionary genetics and genomics of any organism(s) are encouraged. Examples of areas that will be considered include (but are not limited to): microbial evolutionary genomics, vertebrate population genomics, comparative evolutionary genetics and developmental biology, and human evolutionary genomics. Faculty hired through this initiative will participate in the development and teaching of innovative courses. We expect to make two faculty appointments in the 2018-2019 academic year. Openings are at the Assistant Professor level, but exceptional candidates at the Associate Professor and Professor level will be considered.

All applicants should submit: (1) a cover letter with contact information; (2) a CV, (3) description of current and proposed research (3 pages maximum); (4) a statement of teaching experience and interests (1 page maximum); (5) a description of past and/or potential contributions to advancing diversity, inclusion, and equitable access to education (1 page maximum); and (6) up to three reprints and/or preprints. Applicants at the Assistant Professor level should also arrange for three letters of recommendation to be sent on their behalf before the deadline. Applications should be submitted by **October 1, 2018** but later submissions may be considered. Applicants to other searches at the University of Utah with appropriate experience may be considered in this applicant pool, **https://utah.peopleadmin.com/postings/68353**

PRINCETON UNIVERSITY

Princeton Environmental Institute

Princeton Environmental Institute – the interdisciplinary center of environmental research, education and outreach at Princeton University -- seeks distinguished candidates for a **senior appointment (tenured level) in the natural and applied sciences**. The individual should have a demonstrated record of excellence in scholarship and teaching in a field that may include, but is not limited to, global change, biogeochemistry, biodiversity, ecosystem services, and environmental fluids and hydrology. We seek a broad thinker who can integrate perspectives across one or more disciplines including geosciences, ecology and evolutionary biology, chemistry, engineering, hydrology, and/or applied mathematics. The successful candidate will be jointly appointed in the department best suited to her or his research and in the Princeton Environmental Institute. Fields of specialization are open.

Applicants should apply online at **https://www.princeton.edu/acad-positions/position/9341** and should submit a letter of interest along with a vitae, teaching statement, a summary of research plans, and a list of three potential referees. The letter of interest should include the candidate's vision of the field and identify major unanswered questions of interest to the candidate. Evaluation of applicants will begin on **January 7, 2019** and continue until the position is filled.

WISE Women In Science Excel

NWO Nederlandse Organisatie voor Wetenschappelijk Onderzoek

Vacancy Women In Science Excel programme – 2019

Four Women In Science Excel (WISE) positions at five NWO research institutes

The Women In Science Excel (WISE) programme provides talented female researchers with an opportunity to develop or expand their own research group at one of the NWO institutes. NWO believes in providing opportunities for all talented researchers to excel. As such, NWO is committed to improving the gender balance among its research staff. WISE contributes directly by attracting top female researchers and promoting their advancement.

In five recruitment rounds, the NWO institutes will be offering a total of 20 WISE positions. Four WISE positions are being offered in the third round of the programme with the deadline on 4 December 2018 14:00 CET. The NWO research institutes participating in this third round are: ARCNL, CWI, Nikhef, NIOZ and SRON.

We offer a top-level research environment with excellent opportunities to expand and develop your own research programme and personal development plan. A WISE position is for a period of three to five years depending on experience and accomplishments. For more information about the WISE programme and the application procedure, please visit the website: www.nwo.nl/wise.

HFSP

HUMAN FRONTIER SCIENCE PROGRAM (HFSP)

12, quai Saint-Jean, 67080 Strasbourg, France

DIRECTOR OF RESEARCH GRANTS

The International Human Frontier Science Program Organization (HFSPO) seeks to appoint a Director of Research Grants responsible for the administration of the HFSP Research Grant program. The Director of Research Grants must be capable of upholding HFSPO's reputation for undertaking the highest quality peer review with leading scientists from around the world.

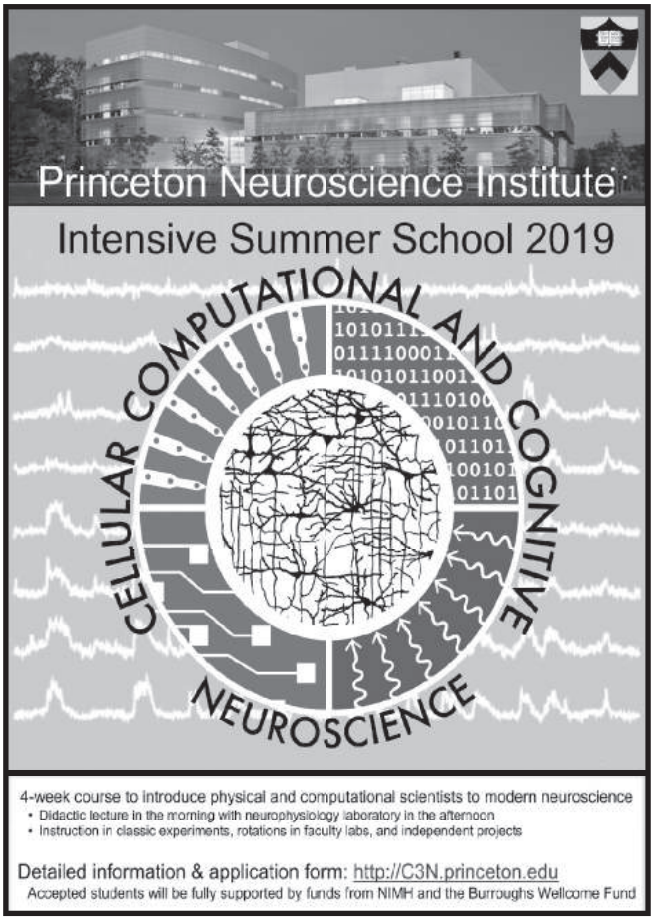
Applications are welcome from all HFSPO Member countries.

Primary job responsibility is to conduct the annual competition and peer review of applications and to oversee administration of awarded grants. The Director is a member of the senior management of the HFSPO, working closely with the Secretary-General and the other Directors of the Secretariat.

The applicants should have a distinguished track record in frontier life science research that is relevant to the interdisciplinary context of the Human Frontier Science Program, a commitment to the mission of HFSPO, and demonstrated organisational skills.

The position is held in Strasbourg, France. Appointments are for 3 years, renewable. A competitive salary will be offered. Removal expenses are available.

Expressions of interest should be sent to **info@hfsp.org** by 31st December 2018. For further information and a full job description see our website at **http://www.hfsp.org/DirectorResearchGrants**.



Stony Brook University

Faculty Searches Biomedical Engineering

The Department of Biomedical Engineering is inviting applications for an Assistant, Associate and/or Full Professor tenure-track faculty position in any BME area. Candidates must have a Ph.D. in BME or related field. A minimum of 2 years of postdoctoral experience is preferred. Outstanding candidates wishing to be considered at the Associate or Full Professor level must have active, funded research programs in their area of expertise. Candidates are expected to develop and maintain competitive extramurally funded interdisciplinary research programs and to excel in teaching at both graduate and undergraduate levels - **www.stonybrook.edu/bme/**.

Qualified individuals should submit their full CV, research and teaching statement, and contact information of three references electronically to: **https://academicjobsonline.org/ajo/jobs/12297**. Questions and inquiries should be directed to: Professor Richard Clark, Ph.D. Chair of the Search Committee, **Richard.Clark@stonybrookmedicine.edu**. Applications should be received by December 1, 2018, but the position will remain open until filled.

Preferred start date: August 2019.

For a full position description, or application procedure, visit: **www.stonybrook.edu/jobs (Ref # F-9941-18-10)**

Stony Brook University is an affirmative action/ equal opportunity employer and educator.

POSITIONS OPEN

FACULTY POSITION IN QUANTITATIVE & COMPUTATIONAL BIOLOGY

The Battelle Center for Mathematical Medicine (BCMM) at The Research Institute at Nationwide Children's Hospital and the Department of Pediatrics of The Ohio State University College of Medicine is seeking to fill an open-rank tenure track position. We are looking for candidates who would extend the quantitative and computational technologies of the BCMM in creative ways; who are interested in both basic quantitative research and collaborative clinical research; and who seek a highly collaborative, biomedical research-focused, environment. Candidates are expected to have a Ph.D. in a quantitative or computational field, or a comparable degree with a quantitative research focus. A generous start-up package is available for highly qualified candidates.

The mission of the BCMM is to assemble and support a broad range of mathematical, statistical, and computational experts for the purposes of conducting cutting-edge quantitative research, with the goal of informing and improving clinical care in pediatrics. Nationwide Children's Hospital is one of the largest freestanding children's hospitals in the United States. The Research Institute occupies over 525,000 square feet of dedicated space, with outstanding shared facilities and core laboratories. Joint faculty appointments in graduate departments at The Ohio State University are available. For more information, please visit our website at www.mathmed.org. Send correspondence, including curriculum vitae, a brief statement of research interests, and contact information for three references to e-mail: **Gail.White@nationwidechildrens.org**.

It is the policy of Nationwide Children's Hospital to provide equal employment opportunities, without regard to race, color, religion, national origin, sex, age or disability and strive to establish a workforce that is representative of the community, patients, and families that we serve.

By **Kathrine Bjerregaard Nielsen**

Putting my grad school angst to use

My phone rings right when I'm about to leave work, as if the person on the other end has been debating the call all day. On the line is a graduate student. At first they are hesitant to talk, but they loosen up when I assure them their question is reasonable and their dilemma is common. The emotion in their voice makes it clear that just going over university guidelines won't be enough. We talk for about 30 minutes, discussing the details of their research. But I am not the student's supervisor or academic adviser. I am a research integrity officer—and I relate to this anxious voice because, not too long ago, it could have been me.

Six months before my Ph.D. thesis was due, I faced a dilemma. The data I had collected over two and a half years made me seriously doubt my main hypothesis, which was based on previous research in my lab—dutifully carried out but perhaps misinterpreted. In my time as a graduate student, I had always worried about unknowingly breaking research rules, such as being too attached to an idea or overselling my findings. This tendency to ruminate led to many philosophical discussions with my patient, supportive supervisor about how scientific research is—and should be—conducted.

"I recall the thrill of results that may cause you to forget about rules and formalities."

Now, I had to make a decision: Would I continue down the path I had laboriously cleared for myself, even though I suspected my conclusions would be flawed? Or would I embark on an investigation of the uncharted territory my results pointed toward, knowing that I might end up with no meaningful data on which to build my thesis?

After a summer holiday on the beach spent thinking and discussing my doubts with a friend and fellow researcher, I chose the latter. Running out of time seemed preferable to producing flawed conclusions.

I spent weeks in the darkness of our microscopy room, collecting preliminary data to convince my supervisor that our previous interpretation was wrong and a very different conclusion made more sense. I was terrified that nothing would come of it, but I also felt the excitement of discovering something novel. I ended up with a less coherent thesis than I had envisioned, but I was proud that it told an honest story about the way science goes sometimes.

Then it came time to find a job. Long before starting the hunt, I had decided that a research career was not for me. Although I wholly enjoyed my Ph.D. experience, I recognized that my tendency to ruminate made me an inefficient scientist. So I looked for opportunities that would keep me out of the lab but still close to research. While trawling through job listings, I was surprised to come across one in which I would be paid to grapple with the same questions that had haunted me during my Ph.D. The title was something I had never heard of before: research integrity officer.

So, what is a research integrity officer? I get closer to answering that each day. I teach courses about good scientific practice and keep myself informed about ethics policies that may affect our researchers. I answer questions from Ph.D. students and professors alike, providing assurance and guidance as best I can. It helps that my own research experience is fresh in my mind. I recall the thrill of results that may cause you to forget about rules and formalities, as well as the sincere wish to do things right and not undermine the reason we do research in the first place: to get as close to a truth as possible.

The main thing I have had to get my head around is that I cannot know everything. Like the researchers I counsel, I have to keep educating myself, keep an open mind, and be humble. I seek advice from legal officers, more experienced administrators, and—not least—the researchers themselves. In this way, I hope to contribute to a more honest, transparent, and accountable scientific community—just what my anxious graduate student self wanted. ■

Kathrine Bjerregaard Nielsen is a research integrity officer at the Technical University of Denmark in Kongens Lyngby. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER