Lasers that can rip matter from empty space *p. 382*

Earliest modern humans outside Africa *pp. 389 & 456*

How Libratus beat poker professionals *p. 418*

Science

$15
26 JANUARY 2018
www.sciencemag.org

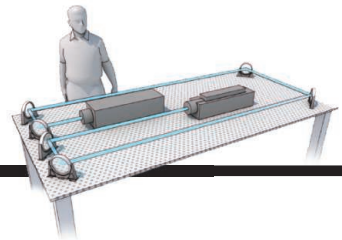
AAAS

NO ROOM TO ROAM

Human alterations of the landscape shorten animal movements *p. 466*

CONTENTS

26 JANUARY 2018 • VOLUME 359 • ISSUE 6374



382
The mightiest lasers

378

NEWS

IN BRIEF

IN DEPTH

FEATURE

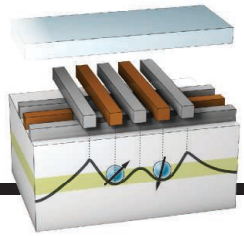
INSIGHTS

PERSPECTIVES

POLICY FORUM

BOOKS *ET AL.*

CONTENTS



393 & 439

Controlling spins in silicon

26 JANUARY 2018 • VOLUME 359 • ISSUE 6374

ON THE COVER

A leopard (*Panthera pardus fusca*) walking through an alley in Mumbai, India. In human-modified landscapes, mammals move distances that are, on average, one-half to one-third the length of their movements in the wild. Reduced travel distances may have far-reaching consequences for species survival and ecosystem functioning. See page 466.
Photo: Nayan Khanolkar/Minden Pictures

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except the 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 and subscription (51 issues): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $1659; 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 #1254 88122. 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:** $23.00 current issue, $28.00 back issue prepaid includes surface postage; bulk rates on request. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

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

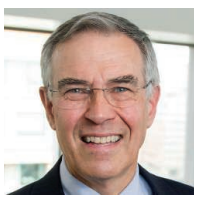
SUBSCRIPTION SERVICES For change of address, missing issues, new orders and renewals, and payment questions: 866-434-AAAS (2227) or 202-326-6417, FAX 202-842-1065. Mailing addresses: AAAS, P.O. Box 96178, Washington, DC 20090-6178 or AAAS Member Services, 1200 New York Avenue, NW, Washington, DC 20005

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

A tale of two cultures



Rush Holt is the chief executive officer of AAAS and executive publisher of the Science *family of journals. rholt@aaas.org*

It is the best of times. It is the worst of times. We are witnessing major advances in almost every field of science, leading to a better understanding of the world and improvements in the quality of people's lives. Yet, scattered distrust of science, neglect of science by public officials, and frequent denial of scientific thinking in many quarters seem to call into question that rosy view of scientific progress. The inconsistency indicates widespread misunderstanding of what science is and how it works. It is up to scientists to fix this.

Examples of exciting advances abound. Take biomedicine, economics, and astrophysics, to choose only three: Immunology is realizing long-dreamed therapies; human behavioral science is increasing the validity of economics; and detection of gravitational waves and concurrent electromagnetic radiation from colliding neutron stars resolves long-standing questions, including how the heavy elements formed. Nevertheless, if science seems remote to nonscientists, and if scientists themselves appear remote and untrustworthy, can the public be counted on to support science into the future? Is there awareness of what would be lost to society if science does not have the public's full endorsement? For example, it is troubling to scientists that in the United States, the president has failed to appoint a science adviser. But even more troubling is that the public has reacted with a yawn.

"...scientists must rebuild public understanding and appreciation of science..."

Several surveys on the perceptions of science point to consistent strong support for science, yet also show very inconsistent underlying views on various scientific subjects, depending on peoples' identification with political party, race and ethnicity, sex, and other group identities [a topic to be covered at next month's annual meeting of the American Association for the Advancement of Science (AAAS), the publisher of *Science*]. This is seen with climate science, vaccinations, genetically modified foods, and other issues. To pick and choose when to believe that the scientific method yields good outcomes suggests that people do not really trust the scientific method. A principle of science is that all findings are provisional. Some seem to think this means science is so uncertain that any opinion or political assertion is as valid as evidence.

Somehow, scientists must rebuild public understanding and appreciation of science and evidence-based thinking. Clearly, it will not be accomplished simply by decrying the lack of trust or failure to appoint science advisers. It must be achieved by demonstrating trustworthiness and the extraordinary effectiveness of science in confronting questions and problems. Scientists must show that evidence-based thinking leads to more reliable policies to create jobs, maintain a healthy environment, or improve teaching. Rather than denouncing the absence of scientists in policy-making positions, the scientific community must raise public understanding to the level where no public official of any party would ever want to be without a science adviser. Scientists must build the recognition that despite occasional errors, and even blunders, scientific thinking has a strong record of success over centuries. Scientists must demonstrate that science and evidence-based thinking are relevant to everyone, and that science is not an arcane practice under the control of a remote, self-interested priesthood.

Science practiced by those who neither make their work accessible to all people, nor make clear their work is for the benefit of all, becomes an impoverished enterprise and risks being unsustainable. It comes down to good science communication—not simply choosing the right words to explain one's research, but actually earning the public's trust that the whole enterprise is intended for societal good. If scientists fail to rebuild the public's understanding and appreciation, this could indeed become the worst of times.

–Rush Holt

10.1126/science.aat0588

> "When even the 'colder' years are rewriting the warmest year record books, we know we have a problem."
>
> **University of Edinburgh climate researcher David Reay**, on the news that 2017 was the second warmest year recorded despite the lack of an El Niño.

IN BRIEF

Edited by **Jeffrey Brainard**

U.S. shutdown ends, for now

Pizza-fueled Senate negotiators failed last week to avert a brief government shutdown.

U.S. scientists must wait at least a few more weeks to learn how much money their favorite agency will receive this year, after Congress ended a 3-day government shutdown, once again delaying a decision on the federal budget. Senate Democrats agreed to end the shutdown after winning a promise from Republican leaders to hold a vote aimed at resolving the status of hundreds of thousands of undocumented immigrants. But what set the stage for the shutdown was Congress's failure to agree on spending levels for the 2018 fiscal year, which began 4 months ago; as a result, the government has been funded through short-term continuing resolutions. Democrats want to increase spending caps equally for domestic and military programs, whereas most Republicans support raising only defense funding. But if the two parties don't strike a spending deal soon, another shutdown is possible when the current continuing resolution ends on 8 February. And if the next one lasts as long as the 16-day shutdown in 2013, it could disrupt a wide range of federally funded research activities, including the vetting of competitive research and training grants and access to unique scientific facilities and assets.

Geoengineering by another name

CLIMATE | Before global warming grows too severe, science agencies worldwide should break out of their current "paralysis" and thoroughly fund research on schemes to intervene in the climate, the American Geophysical Union (AGU), the world's largest geoscience society, said in a statement last week. Two techniques to do this—sucking carbon dioxide out of the atmosphere and injecting particles into the atmosphere to reflect solar rays into space—are among the methods dubbed "geoengineering." That name, however, has helped limit research funding to date, AGU says, because it carries negative connotations or implies excessive confidence in humans' ability to tinker reliably with the climate. From now on, AGU will use a more neutral term: climate intervention.

Indian minister doubts evolution

SCIENCE EDUCATION | A new front has opened in the war on science in India. Last week, India's education minister, Satyapal Singh, took aim at the theory of evolution. Calling himself "a responsible man of science," Singh, a chemist, suggested that Charles Darwin's theory is "scientifically wrong" and "needs to change" in school curricula. India's science academies released a statement endorsed by more than 2000 scientists declaring "it would be a retrograde step to remove the teaching of the theory of evolution from school and college curricula or to dilute this by offering non-scientific explanations or myths." Singh said his ministry intends to hold a conference at which evolutionary theory and creationism "could be debated openly." However, Human Resource Development Minister Prakash Javadekar told the Press Trust of India that the government has no plans "for a national seminar to prove Darwin wrong."

Panel targets sexual harassment

WORKPLACE | Leaders of the U.S. House of Representatives's science panel last week called for an investigation into how federal science agencies police sexual harassment by their grantees. Representative Lamar Smith (R–TX), chairman of the

PHOTO: AARON P. BERNSTEIN/STRINGER/GETTY IMAGES

Published by AAAS

House Committee on Science, Space, and Technology, and its ranking Democrat, Representative Eddie Bernice Johnson (TX), asked the Government Accountability Office (GAO) on 18 January to probe how agencies are enforcing Title IX, the federal law that bars educational institutions receiving taxpayer funds from discriminating based on gender. The agencies are the National Institutes of Health, the National Science Foundation, the departments of agriculture and energy, and NASA. Among other things, the letter asks GAO to report on how many complaints of sexual harassment or assault each agency is handling, and how many have been investigated and resolved since 2013.

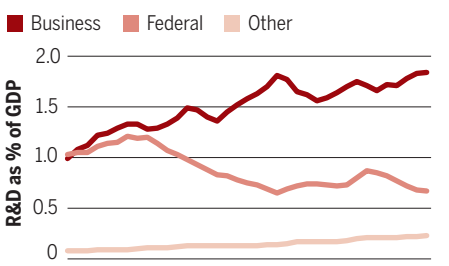
NIH revises clinical trial policy

CLINICAL RESEARCH | Behavioral researchers conducting basic research are bracing for rules from the U.S. National Institutes of Health (NIH), taking effect this week, that reclassify some of their studies as clinical trials. Many maintain that the policy makes no sense and will hinder these studies. Scientists who use tools such as MRI scans to study how the normal brain works, for example, may now be required to register their studies with clinicaltrials.gov, the federal trials database, even though they are not testing treatments. NIH officials have recently eased some concerns by adjusting the rules. Pilot studies with a few volunteers may be exempt. Basic science grants will likely go to the same peer-review panels as before, not ones focused on clinical trials.

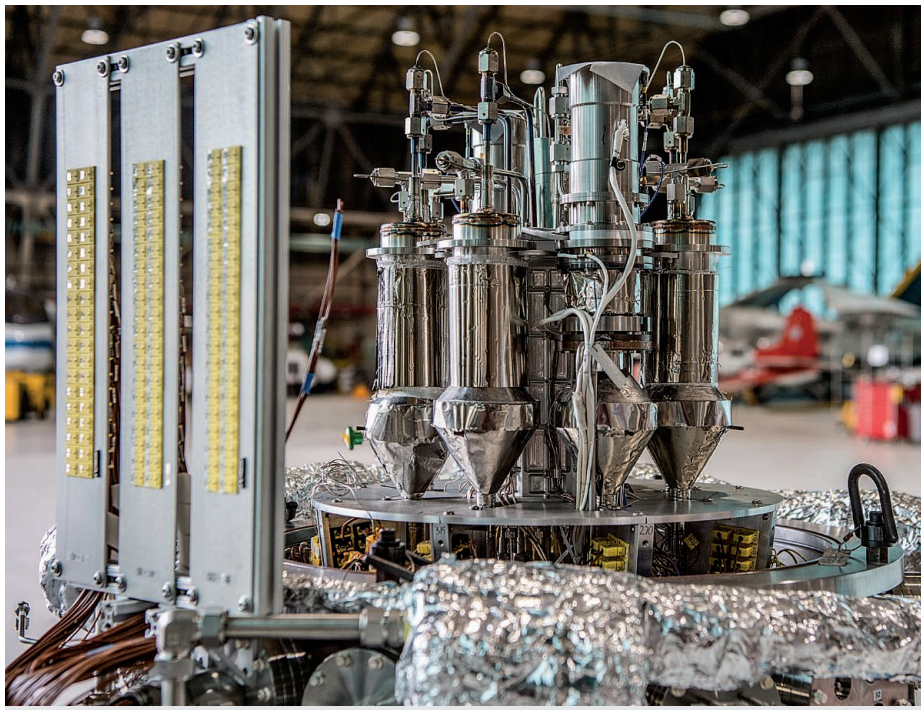
Industry R&D dollars rise

RESEARCH SPENDING | The latest volume of the U.S. National Science Foundation's indispensable guide to science statistics documents the continued erosion of the U.S. government's share of the country's overall investment in research. The balance between business and federal spending has swung from parity in 1980 to a nearly three-to-one margin favoring industry in 2015, the latest year covered by the new *Science & Engineering Indicators* report issued last week. Overall U.S. research spending keeps growing, to $495 billion, representing roughly 2.7% of the country's gross domestic product. But academic leaders feel the U.S. government isn't pulling its weight. Universities are kicking more into the pot: Their share, plus that from other nonprofit organizations, has tripled since 1980, although it's still only one-third the size of the federal contribution.

U.S. spending on R&D as a percentage of gross domestic product (GDP)



CREDITS: (PHOTO) NASA GLENN RESEARCH CENTER; (GRAPHIC) J. YOU/*SCIENCE*; (DATA) *SCIENCE & ENGINEERING INDICATORS*, 2018, NATIONAL SCIENCE FOUNDATION



This fission reactor could generate power in places on Mars or the moon where solar energy is limited.

NASA tests fission power for space missions

NASA is going nuclear. After decades of stalled and canceled projects, NASA this week began its first live, ground-based tests of a nuclear reactor intended for use in deep space or on the surface of the moon or Mars. The Kilopower Reactor Using Stirling Technology (KRUSTY) project is a simple fission reactor, powered by an enriched uranium-235 core the size of a paper towel roll, that would provide up to 10 kilowatts (kW) of electric power. (For comparison, a toaster runs on 1 kW of power.) NASA says launching KRUSTY is less hazardous than sending up the plutonium-238 sources it has used for missions that require less power, such as the Mars Curiosity rover. Unlike those sources, which generate power from radioactive decay, KRUSTY reactors are not highly radioactive until fission is triggered in space. The agency envisions using them to mine water on Mars or the moon for astronauts in regions where solar power is limited.

University returns tobacco grant

TOBACCO RESEARCH | After intense public pressure from the Dutch Cancer Society and other health organizations, Utrecht University (UU) in the Netherlands will return a €360,000 research grant from Philip Morris International (PMI), a leading manufacturer of tobacco products, including Marlboro brand cigarettes. John Vervaele, an economic and European criminal law researcher at UU, had obtained the money from a fund named PMI Impact to study cigarette smuggling in the European Union. The university initially approved the grant because Vervaele had complete academic independence, but changed its mind "in view of the social debate and at the express request of partners who are extremely important to us," UU President Anton Pijpers said in a 17 January statement. UU will now pay for the study, set to begin in February, itself. It was the right decision, says tobacco control expert Michael Siegel of Boston University. Accepting tobacco grants makes universities "pawns" in a public relations strategy, Siegel contends, even if the funds come with no strings attached. PMI

Impact has so far provided about €23 million to 31 other projects at universities, companies, and law enforcement organizations, most of them in Europe.

Thrifty launcher for CubeSats

SPACE EXPLORATION | From a pad in New Zealand, the U.S. company Rocket Lab on 21 January sent its first Electron rocket into orbit, demonstrating a dedicated launch capability for small satellites. For less than $6 million, the firm says, each of its 17-meter-tall rockets can place 150 kilograms into orbit 500 kilometers up—enough capacity to ferry dozens of CubeSats, the toaster-size satellites beloved by space scientists and startups. Rocket Lab hopes to fill a niche for clients wanting fast-paced launches and custom orbits. After the rocket reached orbit, it released three CubeSats made by startups that aim to sell images and weather data. NASA is among Rocket Lab's future customers.

Europe yanks CRISPR patent

BIOTECHNOLOGY | A decision from the European Patent Office (EPO) has put the Broad Institute in Cambridge, Massachusetts, on shaky ground with its intellectual property claims to the gene-editing tool CRISPR. In the United States, the Broad has so far prevailed in a high-profile patent dispute with the University of California (UC), Berkeley. But EPO last week revoked one of its basic patents on the technology. EPO determined that the Broad could not refer to earlier U.S. patent filings to show that its claims predate competing publications and patents from UC and other groups. The rationale: Those earlier filings list an inventor, microbiologist Luciano Marraffini of The Rockefeller University in New York City, who was excluded from the European application. The same issue could jeopardize several other Broad patents that face opposition in Europe. The institute says it plans to appeal the decision.

PHOTO: QIANG SUN AND MU-MING POO/CHINESE ACADEMY OF SCIENCES

Healthy baby monkeys born from cloning

A Chinese group has produced two genetically identical long-tailed macaques using somatic cell nuclear transfer (SCNT), the same technique that resulted in the first cloned mammal, Dolly the sheep, in 1996. The researchers, at the Chinese Academy of Sciences's Institute of Neuroscience in Shanghai, claim the advance allows creating genetically uniform monkeys for biomedical research. Previous attempts to clone monkeys through SCNT produced embryos, but not live healthy babies. A cloned rhesus monkey created through a simpler method, embryo splitting, was born in 1999. To reprogram the cells used in cloning, the Chinese group used a set of enzymes that hadn't been tried on primates before. The success rate was low: Just two healthy baby macaques were born from more than 60 surrogate mothers. And the researchers succeeded only with cells from fetal monkeys; all attempts using adult cells failed. But the group believes that the procedure can be improved. The newborns were named Zhong Zhong and Hua Hua after a Mandarin term for the Chinese nation and people. The team reports its results online this week in *Cell*.

Pharma's resistance fighters

DRUG DEVELOPMENT | Among big pharma companies, GlaxoSmithKline and Johnson & Johnson are making the greatest efforts to fight antimicrobial resistance, according to a ranking revealed 23 January at the World Economic Forum in Davos, Switzerland. The Access to Medicine Foundation in Amsterdam ranked 30 companies based on self-reported information about R&D of new antimicrobials, access to each company's existing antimicrobials by populations who need them the most, and prevention of misuse and overuse of those drugs. Mylan, Cipla, and Fresenius Kabi did best among manufacturers of generic antimicrobial drugs. The report says more work needs to be done to combat resistance; generics companies, for instance, tend to care little about the way their products are used. The report identified 28 new antibiotics now in clinical trials, which sounds impressive but is not enough given the magnitude of the resistance problem, says Jayasree Iyer, director of the Access to Medicine Foundation.

NAS urges lower alcohol limit

PUBLIC POLICY | U.S. states should lower the minimum blood alcohol concentration (BAC) for criminally impaired driving from 0.08% to 0.05% because evidence suggests a lower threshold reduces alcohol-impaired driving injuries and deaths, the U.S. National Academy of Sciences (NAS) recommended in a report last week. The NAS authors write that laboratory and epidemiological studies show that "an individual's ability to operate a motor vehicle … begins to deteriorate at BAC levels well below 0.05%." They cite a 2017 meta-analysis published in *Alcoholism: Clinical & Experimental Research*. Analyzing 11 studies in six non-U.S. countries and the state of Maine in the 1980s, its authors estimated that dropping the limit to 0.05% or lower resulted in an 11.1% decline in fatal alcohol-related crashes. They also predicted that lowering the legal limit to 0.05% in all 50 states would save 1790 lives each year. Critics accused NAS of overreach, arguing that a 0.05% limit would ensnare tens of thousands of unimpaired drivers.

IN DEPTH

Edward Jenner's smallpox vaccine was based on horsepox, claim scientists who synthesized the virus.

BIOSECURITY

Critics see only risks, no benefits in horsepox paper

Scientists say their labmade virus could make a new smallpox vaccine. Others call it a "mistake" and a "stunt"

By **Kai Kupferschmidt**

On 19 January, a highly controversial study in which researchers synthesized a smallpox relative from scratch finally saw the light of day. The paper, published in *PLOS ONE*, spells out how virologist David Evans at the University of Alberta in Edmonton, Canada, and his research associate Ryan Noyce ordered bits of horsepox DNA from the internet, painstakingly assembled them, and packaged them into viral particles that were able to infect cells and reproduce.

The study stirred alarm when *Science* first reported it in July 2017 because it might give would-be terrorists a recipe to construct smallpox virus, a major human scourge vanquished in 1980. And now that it's out, many scientists say the paper doesn't answer the most pressing question: Why did they do it?

The team claims its work, funded by Tonix, a pharmaceutical company headquartered in New York City, could lead to a safer, more effective vaccine against smallpox. But safe smallpox vaccines already exist, and there appears to be no market for a horsepox-based replacement, says virologist Stephan Becker of the University of Marburg in Germany. "It simply does not add up," Becker says. Given the apparent lack of benefits, publishing the paper was "a serious mistake," says Thomas Inglesby, director of the Center for Health Security at the Johns Hopkins University Bloomberg School of Public Health in Baltimore, Maryland. "The world is now more vulnerable to smallpox." "It almost feels like a stunt in search of justification," adds microbiologist David Relman of Stanford University in Palo Alto, California, who says he personally asked *PLOS ONE* Editor-in-Chief Joerg Heber not to publish the paper.

A spokesperson for the journal's publisher, PLOS, wrote in an email that a committee on "dual use research of concern" at *PLOS ONE* unanimously agreed that the benefits of publication, including "the potential improvements in vaccine development," outweighed the risks. But Inglesby says that given its implications, national and global health authorities should have approved the work; currently, there is no requirement for such a review. "This ought to be a wake-up call for science agencies and governments," Inglesby says.

Vaccination against smallpox ended worldwide in the 1980s, and most people have no immunity. Experts worry that the virus that causes it, variola, could be used in an act of bioterror or biological warfare. The last variola samples are kept under tight security in Russia and the United States. But the paper could provide another route for terrorists.

World Health Organization recommendations ban the synthesis of variola's full genome, and ordering its DNA might be difficult because some synthesis companies screen their orders. "This is not something that can be done tomorrow in any lab," says Gregory Koblentz, a biodefense expert at George Mason University in Arlington, Virginia. (The horsepox virus itself, thought to have gone extinct in nature, is not known to cause human disease.)

But over time, other labs will adopt the technique used to make it and gain the ability to recreate smallpox as well, Koblentz says, "and that just creates a huge vulnerability." "If anyone wants to recreate another poxvirus, they now have the instructions to do that in one place," says Andreas Nitsche of the Robert Koch Institute in Berlin.

The work arose in part from Tonix CEO Seth Lederman's fascination with Edward Jenner, the English physician who invented the first smallpox vaccine in 1796. Lederman, who says he's "obsessed" with Jenner and has been working on a biography of the scientist for 20 years, is convinced that Jenner's original vaccine, now named vaccinia, was derived from a virus circulating in horses, not cows, as folklore has it. "Jenner essentially says that" in his most famous paper, Lederman says. A genetic analysis published in October 2017 strengthened his case.

In a conversation about one of Evans's earlier vaccinia papers, Lederman told Evans that he'd love to know whether horsepox still works as a vaccine, but that he had had trouble getting hold of the virus. The only known samples are stored at the U.S. Centers for Disease Control and Prevention, which would not give permission for them to be used commercially, Lederman says.

"I think if you're really interested in that, I know how to make it," Evans says he responded. Tonix provided about $100,000 for the project, and Lederman became a coauthor on the paper.

Vaccinia can cause a serious rash and inflammation of the brain and the heart muscle; it's especially dangerous for immunocompromised people. It has been estimated to kill between 1.4 and 8.4 people per million people who are vaccinated. Lederman believes that may be because the virus used by Jenner evolved during the almost 180 years it was used, racking up mutations that made

IMAGE: COLLECTION OF THE UNIVERSITY OF MICHIGAN HEALTH SYSTEM. GIFT OF PFIZER INC., UMHS.23

it better at replicating in humans, which led to more serious side effects. A virus closer to Jenner's vaccine might be safer and more effective, he argues.

But safer vaccines are already on the market. Bavarian Nordic, a German-Danish company, produces a weakened vaccinia strain called Modified Vaccinia Ankara (MVA) that doesn't replicate in humans. It was originally developed by the Bavarian government and given to 150,000 German children in the 1970s; it has proved safe in more recent trials in HIV-positive people and stem-cell transplant patients. The European Union and Canada have licensed MVA, and U.S. regulators are expected to follow suit this year. The U.S. government has already stockpiled 28 million doses and signed an option for 13 million additional doses in September 2017. U.S. government agencies "have been signaling for many years that we are not going to invest in a whole new smallpox vaccine," Koblentz says.

In the paper, Evans and his co-authors don't discuss MVA, or a similar vaccine developed in Japan named LC16m8, which "makes no sense to me," Becker says. But Lederman argues that MVA's efficacy was never properly tested because it was developed when smallpox was already very rare. Using it in a real outbreak would be "taking a huge uneducated guess with the health of an entire country's population," he says. "We believe there is more historical evidence supporting the efficacy of horsepox or horsepox-like vaccine." Tonix has shown that horsepox can protect mice from an otherwise lethal dose of vaccinia, he says, and is investing in production of the virus under the strict conditions needed for human trials.

Whether the world really needs another smallpox vaccine is a crucial question, says Diane DiEuliis, a biosecurity consultant at the National Defense University in Washington, D.C. "I would have liked to see an open debate about that at the outset of these experiments," she says

Evans, for his part, says he "was far more interested in advancing the technology" than in developing a vaccine. Creating poxviruses from scratch could help answer basic questions about their biology, he says. Besides, "It's a very powerful technology and I do see there being lots of applications for it," for instance, in the design of cancer-fighting viruses.

Lederman and Evans disagree that the publication makes the world less secure. The work could aid preparedness by showing that synthetic smallpox, long a theoretical concern, is a real possibility, they say. On that point, at least, Nitsche agrees. "For that reason alone, it's good someone has finally done this." ■

The SpaceX Falcon Heavy is the most powerful rocket since the Saturn V that took humans to the moon.

PHOTO: SPACEX/FLICKR

SPACE

Heavy-lift rocket poised to boost space science

Falcon Heavy could enable faster, heavier, cheaper missions

By **Adam Mann**

SpaceX CEO Elon Musk has said he'll consider it a win if his enormous new Falcon Heavy rocket even escapes the launch pad. This week, the rocket was set to fire its engines in a test at the Kennedy Space Center in Florida, which would clear the way for an inaugural launch in the coming weeks. Space scientists will be rooting for it, too. With its heavy-lift capability, the rocket can fling larger probes to distant planets more quickly—and, perhaps, more cheaply—than previous rockets.

"We can think about follow-up missions across the outer solar system, Mars sample return, even missions to Venus or Mercury," says planetary scientist Alan Stern of the Southwest Research Institute in Boulder, Colorado, who was also an independent consultant for SpaceX between 2010 and 2012.

Created by strapping together three of the company's Falcon 9 rockets, the Falcon Heavy is 70 meters tall, the most powerful rocket since the Saturn V that took humans to the moon. It is expected to carry up to 64,000 kilograms to low-Earth orbit, more than twice the payload of the biggest currently available vehicle, United Launch Alliance's (ULA's) Delta IV Heavy. Moreover, the new rocket's booster stages can be reused, which SpaceX claims will save money. It says a Falcon Heavy launch will start at a mere $90 million—less than 20% of the Delta IV Heavy's cost.

Such price tags could transform mission planning for NASA and other space agencies, Stern says. "You're talking about savings of hundreds of millions of dollars, which is sufficient to create whole new missions just from the savings." Of course, were NASA to save on launches, Congress could take that money and use it elsewhere, says Henry Hertzfeld, who studies space policy at The George Washington University in Washington, D.C. He adds that the launch fees that government agencies pay tend to be negotiated in long-term contracts, based on payload needs, and don't necessarily align with prices published on a company website.

The business case for Falcon Heavy involves taking commercial satellites cheaply to low-Earth orbit, and sending tourists on slingshots around the moon. But researchers have also been penciling the rocket into their plans ever since its design was unveiled in 2011. Studies for NASA's $2 billion Europa Clipper mission have often included the possibility of hitching a ride on one as a money-saving measure. The 3.6-ton probe, slated to launch as early as 2022, would conduct flybys of Jupiter's intriguing frozen moon to map its icy crust and subsurface ocean. Planners still expect it to fly on NASA's own heavy-lift vehicle, the nearly 100-meter-tall Space Launch System (SLS). But the first test launch of the SLS

has slipped to 2020, and each launch is expected to cost about $1 billion.

Other possible targets for Falcon Heavy include Saturn's moons Titan and Enceladus and the ice giants Neptune and Uranus. Stern, who leads a NASA mission that flew past Pluto in 2015, says teams are considering using the rocket to send a probe with enough fuel to slow down and orbit the distant world. SpaceX has said that Falcon Heavy could deliver 2 to 4 tons to the surface of Mars—opening the way to more ambitious missions than the 1-ton Curiosity rover.

Astronomers are also thinking about what heavy lift can do for them. Each component of NASA's upcoming 6.2-ton James Webb Space Telescope, with a 6.5-meter mirror, had to be both lightweight and yet hardy enough to withstand rigorous shaking during launch, two often incompatible requirements. With Falcon Heavy's additional lift, researchers planning the Large UV Optical Infrared Surveyor telescope, a proposed mission for the 2020s with a mirror at least 9 meters across, could focus less on reducing weight and more on delivering a great scientific instrument, says Matt Mountain, president of the Association of Universities for Research in Astronomy in Washington, D.C. "If we don't have to fight for mass, the testing is greatly simplified and you can launch more ambitious systems."

SpaceX's big rocket will face competition in the coming years, and not just from the SLS. Another private company, Blue Origin, intends to debut its reusable New Glenn rocket in 2020, and ULA is working on a vehicle called Vulcan. The competition could lower prices for researchers, says Phil Larson, an aerospace expert at the University of Colorado in Boulder and a former communications director at SpaceX. "You could see not just governments having space programs, but private entities doing more in space, and maybe universities," he says.

Then again, the pool of heavy rockets will remain small for the foreseeable future, and researchers shouldn't expect rock-bottom prices anytime soon, Hertzfeld says. "This is not your Econ 1 definition of competition," he says. "It certainly should help NASA and others with prices, but how much that influences all of this is a study in and of itself."

SpaceX already has customers lining up, including the Saudi Arabian communications firm Arabsat, the London-based mobile services provider Inmarsat, and the U.S. Air Force. Its first payload will be something a bit more personal—Musk's own Tesla Roadster, painted midnight cherry and playing David Bowie's "Space Oddity." ■

Adam Mann is a journalist in Oakland, California.

DATA: AGNES MCMAHON/ENIGMA

NEUROSCIENCE

In thousands of brain scans, group seeks clues to diseases

Amid probe of many brain disorders, ENIGMA finds common structural changes in diverse kinds of epilepsy

By **Giorgia Guglielmi**

ENIGMA, the world's largest brain mapping project, was "born out of frustration," says neuroscientist Paul Thompson of the University of Southern California in Los Angeles. In 2009, he and geneticist Nicholas Martin of the Queensland Institute of Medical Research in Brisbane, Australia, were chafing at the limits of brain imaging studies. The cost of MRI scans limited most efforts to a few dozen subjects—too few to draw robust connections about how brain structure is linked to genetic variations and disease. The answer, they realized over a meal at a Los Angeles shopping mall, was to pool images and genetic data from multiple studies across the world.

After a slow start, the consortium has brought together nearly 900 researchers across 39 countries to analyze brain scans and genetic data on more than 30,000 people. In an accelerating series of publications, ENIGMA's crowdsourcing approach is opening windows on how genes and structure relate in the normal brain—and in disease. This week, for example, an ENIGMA study published in the journal *Brain* compared scans from nearly 4000 people across Europe, the Americas, Asia, and Australia to pinpoint unexpected brain abnormalities associated with common epilepsies. ENIGMA is "an outstanding effort. We should all be doing more of this," says Mohammed Milad, a neuroscientist at the University of Illinois in Chicago who is not a member of the consortium.

ENIGMA's founders crafted the consortium's name—Enhancing NeuroImaging Genetics through Meta-Analysis—so that its acronym would honor U.K. mathematician Alan Turing's code-breaking effort targeting Germany's Enigma cipher machines during World War II. Like Turing's project, ENIGMA aims to crack a mystery. Small brain-scanning studies of twins or close relatives done in the 2000s showed that differences in some cognitive and structural brain measures have a genetic basis. At the same time, researchers performing genome-wide association studies were scanning the DNA of large numbers of people for millions of bases that naturally vary. But efforts to link such genetic variations to differences in brain structure faltered, because imaging studies had too few participants to draw persuasive conclusions. What's more, privacy policies often prevented researchers from sharing brain imaging data.

ENIGMA hoped to solve such problems through a meta-analysis approach, in which all the consortium's research centers are given the same protocols to analyze brain images. Then the results are weighed and combined. Sharing processed rather than raw data means ENIGMA researchers don't have to disclose details about individual patients, and the larger sample sizes enable them to draw "reliable conclusions" on how genetic variation alters brain structure, Thompson says.

In its first project, the consortium, in collaboration with another multicenter network, analyzed the brains scans and DNA of 21,000 people and linked specific genes to brain size and intelligence. For example, people with a specific variant of a gene called *HMGA2* had larger brains and scored on average 1.29 points better on standardized IQ tests.

Speaking to *New Scientist* at the time, Steven Pinker, a neuropsychologist at Harvard Univer-

Crowdsourcing brain images

Born out of a discussion during a meal in a mall, the ENIGMA brain imaging consortium has grown rapidly after a slow start.

	2009	2012	2016	2017	% INCREASE (2016–2017)
Working groups	3	9	20	37	85
Members (roughly)	—	25	500	900	80
Sites (roughly)	—	20	200	230	15
Countries	—	12	33	39	18.18
Projects	—	18	50	166	232
Data sets	—	~8000	~30,000	~53,000	76.67
Publications	—	2	15	36	140

sity, called the result "an important finding, assuming it holds up." But other scientists argued that the effect of *HMGA2* on intelligence was too small to be significant.

Still, the findings, published in 2012 in *Nature Genetics*, gave momentum to ENIGMA's efforts. In a follow-up study of nearly 31,000 people, the researchers examined the size of brain regions involved in memory, movement, learning, and motivation, and looked for associated genetic variants. Five newly identified genetic variants influenced the size of two regions, the putamen and caudate nucleus, the team reported in 2015 in *Nature*.

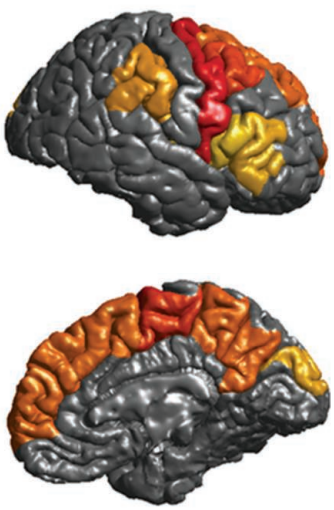
Such structural differences are seen not just in healthy people, but also in people with disease; both the caudate nucleus and the putamen are known to be smaller in Alzheimer's and Parkinson's disease patients, for example. In 2014, the ENIGMA network was awarded a 4-year, $11 million grant from the National Institutes of Health to pursue such correlations. These studies have tended to leave out genetic factors—poorly known for many diseases—and instead bring statistical power to the search for structural differences between the brains of people with and without a disorder.

A recent example is an analysis of schizophrenia patients' white matter, the tissue that carries impulses between neurons and across the brain. After looking at brain images from 1963 schizophrenia patients and 2359 healthy people, an ENIGMA working group reported last year in *Molecular Psychiatry* that the white matter wiring is impaired all over the patients' brains, rather than only in specific areas, as many researchers previously thought.

Also last year, ENIGMA researchers working on posttraumatic stress disorder (PTSD) found that the hippocampus—a region that is key to turning short-term memories into long-term ones—was, on average, smaller in patients than in healthy people. Some smaller studies had hinted at this reduction, which the team reported in *Biological Psychiatry*, but it's now "a definitive finding," says Rajendra Morey, a neuroscientist at Duke University in Durham, North Carolina, who leads the ENIGMA-PTSD working group.

The latest ENIGMA effort found that compared with healthy individuals, people with epilepsy tended to have areas of reduced thickness in the cerebral cortex, which influences cognition, memory, and consciousness. They also had a smaller right thalamus, a brain structure that relays information to and from the cerebral cortex. Some of these features are limited to specific types of epilepsy, the researchers found, but others appear in many epilepsies, not just the most severe. That finding, says Andrew Bagshaw, a neuroscientist at the University of Birmingham in the United Kingdom who is not part of the consortium, "starts to give a very different picture of epilepsies."

"These changes are probably independent of the actual causes, mostly unknown," says Sanjay Sisodiya, a neuroscientist at University College London, who led the research. The ENIGMA-epilepsy working group hopes to pursue those causes in future studies, which might track brain changes in people over time. "All ENIGMA projects start off with one relatively straightforward study and then expand as confidence grows and more people join the consortium," Sisodiya says.



People with epilepsy tend to have areas of thinning (colored) in the cerebral cortex, the latest ENIGMA study has found. The darker the color, the thinner the cortex.

Milad considers the effort "impressive," but says that one limitation of the study, and of most work from ENIGMA, is that the brain images come from a variety of MRI scanners and therefore might not be uniform. The challenge may grow in the future, as ENIGMA adds data on brain activity, from functional MRI or electroencephalography, to the brain mapping. By showing the brain at work, the new data should help researchers link structural differences to function, but they could prove even more difficult to standardize and weigh, Milad says.

Emily Jones, a neuroscientist at Birkbeck University in London, echoes this concern but adds that the large sample sizes available to ENIGMA researchers should help compensate for inconsistencies between labs.

Geneticist and ENIGMA Co-Founder Barbara Franke of Radboud University in Nijmegen, the Netherlands, expects that as the consortium combines genetics with data from ever larger numbers of brains, it will start to trace the pathways leading from genes to brain alterations to diseases. Ultimately, diagnosis and prevention will benefit, Franke predicts. However, she cautions, "We're still a long way from it." ■

SCIENTIFIC COMMUNITY

Crackdown threatens science in Turkey

Verdict expected soon in trial of dual-nationality scientist accused of terrorism

By **Kristen McTighe**

Early next month, Serkan Golge will learn whether his nightmare will continue. In July 2016, he and his family were in southern Turkey wrapping up a visit to relatives and preparing to return home to Houston, Texas, where the Turkish-American space scientist studies the effects of radiation on astronauts. But before his wife, Kubra, could finish strapping their two young sons into car seats, police arrived and took Golge away.

The police accused him of spying for the Central Intelligence Agency, and last April they charged the dual citizen with terrorism. "We were all stunned. It didn't make any sense," says computer scientist Alicia Hofler, a former colleague of Golge's at Old Dominion University in Norfolk, Virginia. A verdict is expected on 8 February; if convicted, Golge faces up to 15 years in prison.

He is one of several U.S. citizens and thousands of Turkish academics caught up in a crackdown following a July 2016 coup attempt. Scores of scientists are in prison, and many more have lost their jobs. Most academics now need permission to travel abroad. Turkey's president, Recep Tayyip Erdoğan, asserts the measures are necessary for national security: to root out allies or sympathizers of Fethullah Gülen, a cleric living in exile in Pennsylvania whom Erdoğan has accused of masterminding the coup.

Harsh measures under an ongoing state of emergency could threaten the future of Turkish science, observers say. As long as a travel ban persists and "scientists currently jailed are not released, people won't be able to maintain their relations abroad," says Eugene Chudnovsky, a physicist at the City University of New York's Lehman College and co-chair of the Committee of Concerned Scientists, a nonprofit dedicated to protecting scientists' human rights. Ironi-

IMAGES: C. WHELAN *ET AL*., *BRAIN* 10.1093/BRAIN/AWX341

Under a crackdown led by Turkey's president, Recep Tayyip Erdoğan, scores of scientists have been imprisoned and many more have lost their jobs.

PHOTO: CHRIS MCGRATH/GETTY IMAGES

cally, before the coup Erdoğan was widely regarded as a patron of science, Chudnovsky says. "Many Turkish scientists will say that their situation had improved tremendously in terms of their economic situation, their ability to travel internationally," he says. "Now, of course, scientists are worried."

Soon after the failed coup, authorities raided the Scientific and Technological Research Council of Turkey, the government research funding agency known by its Turkish acronym TÜBİTAK, arresting dozens. The purge reached the higher education system, where some 5000 academics accused of ties to Gülen have been dismissed, suspended, or forced to resign (*Science*, 16 September 2016, p. 1192).

That frenzy ensnared Golge. According to Kubra, an estranged family member who held a grudge over an inheritance dispute told authorities Golge was a terrorist and a spy. He has denied the charges; in court months later, the relative who tipped off police stated he was only 1% sure the accusations he leveled were actually true, says Kubra, who has attended the trial. As evidence of Golge's guilt, prosecutors pointed to the fact that he, like many Turks, holds an account in a bank owned by Gülen followers, and he studied at a university with ties to Gülen.

Another Turkish-American scientist caught in the dragnet is Ismail Kul, a chemist at Widener University in Chester, Pennsylvania. According to Turkish media reports, Kul and his brother were arrested on a visit to Turkey in 2016 and accused of participating in the attempted coup. In court, Kul denied the charges, though he acknowledged having met Gülen several times since 2010, when he was introduced to the cleric by a Turkish legislator in Erdoğan's ruling Justice and Development (AK) Party. (AK and the Gülen movement were allies with similar ideologies until a conflict between the two erupted in 2013.) Kul is out on bail while his trial continues.

Those Turkish scientists spared persecution have seen international collaborations wither. Academics now must obtain permission from university administrators to travel abroad, and TÜBİTAK has reportedly curtailed travel scholarships for students and researchers. Many who manage to get out have not come back, accelerating Turkey's brain drain, according to İlker Birbil, a Turkish data scientist at Erasmus University in Rotterdam, the Netherlands.

Birbil, who left Turkey in January 2017, faced difficulties after signing a petition in 2016 calling on the government to resume peace talks with Kurdish militants. Taking a stand had repercussions, he says. "We suddenly realized that it was impossible for us to get funding from TÜBİTAK," he says. "In a nutshell, they don't really go after merit anymore." After signing the petition, Izge Günal says he was forced to resign from the orthopedics department at Dokuz Eylül University in İzmir, Turkey. Günal, who remains in Turkey, foresees a steady decline of Turkish science, "replacing Enlightenment thought with superstitions." In public comments, Turkey's science minister, Faruk Özlü, has denied that the government is pressuring scholars or interfering in TÜBİTAK's reviews.

What happens to Kul and Golge could hinge on tensions between Turkey and the United States, which has supported Kurdish rebels in Syria and has balked at extraditing Gülen. In a televised speech last September, Erdoğan said Andrew Brunson, a pastor and U.S. citizen jailed in Turkey, would not be returned until the United States extradited Gülen. The comments left Golge's family fearful that he, too, could effectively be used as a bargaining chip.

The uncertainty is excruciating, Kubra says. For the first 3 months after her husband was arrested, "Every cell in my body was aching because of the pain in my soul," she says. "My life turned upside down and I felt like my family was falling apart."

Hers is only one of many families in limbo. ■

Kristen McTighe is a freelance journalist in Istanbul, Turkey.

Melbourne, Australia, is a hotbed for thunderstorm asthma, with seven outbreaks since 1984.

METEOROLOGY

Australian state forecasts deadly thunderstorm asthma

Scientists probe ways that storms weaponize pollen

By **Katherine Kornei**

On 21 November 2016, an evening thunderstorm rumbled across Victoria state in Australia—a normal springtime event with abnormal consequences. "We've seen a rise in breathing probs tonight following the weather," Ambulance Victoria tweeted at 8:40 p.m. That would turn out to be an understatement: Thousands of people experiencing respiratory distress were rushed to hospitals, and at least nine died. Thunderstorm asthma had struck again.

First identified in the 1980s, thunderstorm asthma is a phenomenon that meteorologists are just beginning to understand. It is known to require high airborne pollen counts, along with thunderstorm winds to sweep up pollen grains and send them pouring down onto a vulnerable populace, says Andrew Grundstein, a climate scientist at the University of Georgia in Athens. He is now sharpening that picture. Earlier this month, at a meeting of the American Meteorological Society in Austin, he presented a study of seven thunderstorm asthma events in Melbourne, the capital of Victoria and the epicenter of this odd confluence of meteorology and medicine. He found a common pattern: All seven featured high pollen levels, multiple storm cells, and strong but not damaging winds.

Late last year, Victoria unveiled a forecasting system that gives residents 3 days' warning of a likely thunderstorm asthma event. The forecast is based on pollen levels and the likely strength of wind gusts in storms, says Jeremy Silver, an atmospheric scientist at the University of Melbourne working with Grundstein. But it's not a perfect system. "This is a bit rudimentary," Silver says. "There are going to be false positives." By determining the types of thunderstorms that have been dangerous in the past, Grundstein and Silver hope to improve predictions. "The work we're doing can help to refine that warning system," Grundstein says.

Thunderstorm asthma events have been reported across North America, Europe, and the Middle East. But Grundstein says the ingredients for severe incidents come together in Melbourne: nearby fields of pollen-producing rye grass, regular thunderstorms, and a large population. The rising convective updrafts that power thunderstorms sweep up pollen grains and send them kilometers into the sky, where they encounter high humidity levels. Scientists believe that when a grain absorbs moisture, it can swell and rupture into hundreds of smaller pieces. This pollen shrapnel is much more dangerous to the lungs than whole pollen grains, says Guy Marks, a respiratory physician at The University of New South Wales in Sydney, Australia, who studies thunderstorm asthma. "They're small enough to enter the lower respiratory tract and cause an inflammatory response." And in the right kind of thunderstorm, these fragments ride cooler downdrafts that crash into Earth's surface and create lateral gusts, which scatter the pollen for tens of kilometers.

Grundstein and his collaborators mined meteorological records to look for commonalities among the seven Melbourne thunderstorm asthma events since 1984. All took place in November, when pollen levels were already high—sometimes more than 100 grains per cubic meter of air. Multiple storm cells were typically involved. "More thunderstorms spread more pollen around," Grundstein says. And the gusty lateral winds needed to be strong, but not necessarily extreme. "Damaging winds aren't necessary," Grundstein says. "Pollen is light."

Last year, Victoria's government put $15.6 million toward developing its warning system, which went live on 1 October 2017, just in time for the start of the grass pollen season. It combines thunderstorm forecasts provided by Australia's Bureau of Meteorology with pollen counts coordinated by Ed Newbigin, a botanist at the University of Melbourne. The pollen monitoring system is based on eight counting stations across the state, powered by antique air pumps—"1950s British technology," he says—that direct air over adhesive-coated glass slides. "We're literally counting pollen grains," Newbigin says. Forecasters combine these meticulous observations with temperature, humidity, wind, and satellite-derived grass coverage data to predict future pollen levels.

The system's advance warning will help people prepare for events by giving them time to collect asthma medications and make plans to stay indoors. But refining Victoria's forecasts requires a better understanding of the 3D structure of these storms, says Silver, a challenge because thunderstorm asthma events occur just a few times per decade. Grundstein isn't deterred, however: He's planning a multiyear study of thunderstorm asthma in the United States. ■

Katherine Kornei is a journalist in Portland, Oregon.

PHOTO: ISTOCK.COM/SW_PHOTO

PARTICLE PHYSICS

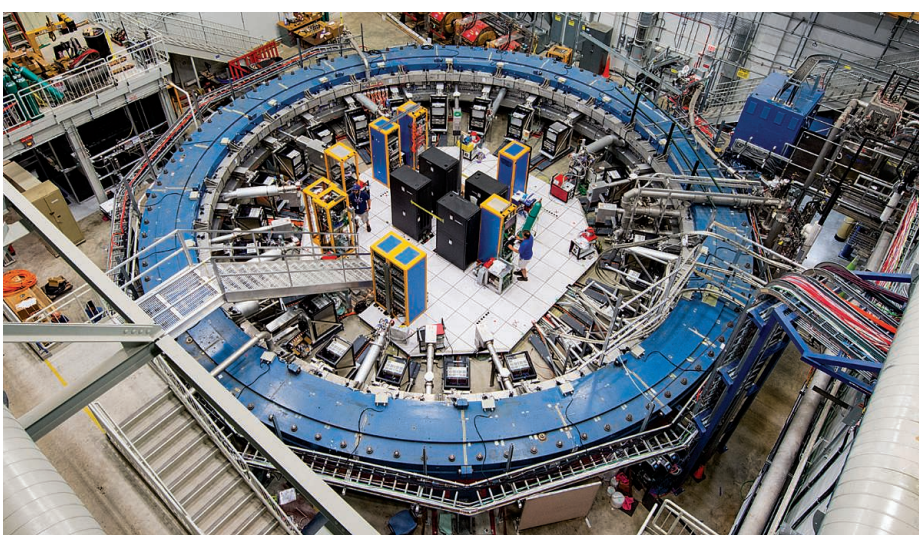
Muon's magnetism could point to new physics

After a hiatus of nearly 20 years, experimental scrutiny of fleeting particle resumes

By **Adrian Cho**

Next week, physicists will pick up an old quest for new physics. A team of 190 researchers at Fermi National Accelerator Laboratory (Fermilab) in Batavia, Illinois, will begin measuring to exquisite precision the magnetism of a fleeting particle called the muon. They hope to firm up tantalizing hints from an earlier incarnation of the experiment, which suggested that the particle is ever so slightly more magnetic than predicted by the prevailing standard model of particle physics. That would give researchers something they have desired for decades: proof of physics beyond the standard model.

"Physics could use a little shot of love from nature right now," says David Hertzog, a physicist at the University of Washington in Seattle and co-spokesperson for the experiment, which is known as Muon g-2 (pronounced "gee minus two"). Physicists are feeling increasingly stymied because the world's biggest atom smasher, the Large Hadron Collider (LHC) near Geneva, Switzerland, has yet to blast out particles beyond those in the standard model. However, g-2 could provide indirect evidence of particles too heavy to be produced by the LHC.



The magnetism of muons is measured as the short-lived particles circulate in a 700-ton ring.

PHOTO: FERMILAB

The muon is a heavier, unstable cousin of the electron. Because it is charged, it will circle in a magnetic field. Each muon is also magnetized like a miniature bar magnet. Place a muon in a magnetic field perpendicular to the orientation of its magnetization, and its magnetic polarity will turn, or precess, just like a twirling compass needle.

At first glance, theory predicts that in a magnetic field a muon's magnetism should precess at the same rate as the particle itself circulates, so that if it starts out polarized in the direction it's flying, it will remain locked that way throughout its orbit. Thanks to quantum uncertainty, however, the muon continually emits and reabsorbs other particles. That haze of particles popping in and out of existence increases the muon's magnetism and make it precess slightly faster than it circulates.

Because the muon can emit and reabsorb any particle, its magnetism tallies all possible particles—even new ones too massive for the LHC to make. Other charged particles could also sample this unseen zoo, says Aida El-Khadra, a theorist at the University of Illinois in Urbana. But, she adds, "The muon hits the sweet spot of being light enough to be long-lived and heavy enough to be sensitive to new physics."

From 1997 to 2001, researchers on the original g-2 experiment at Brookhaven National Laboratory in Upton, New York, tested this promise by shooting the particles by the thousands into a ring-shaped vacuum chamber 45 meters in diameter, sandwiched between superconducting magnets. Over hundreds of microseconds, the positively charged muons decay into positrons, which tend to be spat out in the direction of the muons' polarization. Physicists can track the muons' precession by watching for positrons with detectors lining the edge of the ring.

The g-2 team first reported a slight excess in the muon's magnetism in 2001. That result quickly faded as theorists found a simple math mistake in the standard model prediction (*Science*, 21 December 2001, p. 2449). Still, by the time the team reported on the last of its Brookhaven data in 2004, the discrepancy had re-emerged. Since then, the result has grown, as theorists improved their standard model calculations. They had struggled to account for the process in which the muon emits and reabsorbs particles called hadrons, says Michel Davier, a theorist at the University of Paris-South in Orsay, France. By using data from electron-positron colliders, he says, the theorists managed to reduce this largest uncertainty.

Physicists measure the strength of signals in multiples of the experimental uncertainty, σ, and the discrepancy now stands at 3.5 σ—short of the 5 σ needed to claim a discovery, but interesting enough to warrant trying again.

In 2013, the g-2 team lugged the experiment on a 5000-kilometer odyssey from Brookhaven to Fermilab, taking the ring by barge around the U.S. eastern seaboard and up the Mississippi River (*Science*, 14 June 2013, p. 1277). Since then, they have made the magnetic field three times more uniform, and at Fermilab, they can generate far purer muon beams. "It's really a whole new experiment," says Lee Roberts, a g-2 physicist at Boston University. "Everything is better."

Over 3 years, the team aims to collect 21 times more data than during its time at Brookhaven, Roberts says. By next year, Hertzog says, the team hopes to have enough data for a first result, which could push the discrepancy above 5 σ.

Will the muon end up being a portal to new physics? JoAnne Hewett, a theorist at SLAC National Accelerator Laboratory in Menlo Park, California, hesitates to wager. "In my physics lifetime, every 3-σ deviation from the standard model has gone away," she says. "If it weren't for that baggage, I'd be cautiously optimistic." ■

FEATURES

THE LIGHT FANTASTIC

Physicists in China and elsewhere are vying to build lasers so powerful they could rip apart empty space

By **Edwin Cartlidge**

Inside a cramped laboratory in Shanghai, China, physicist Ruxin Li and colleagues are breaking records with the most powerful pulses of light the world has ever seen. At the heart of their laser, called the Shanghai Superintense Ultrafast Laser Facility (SULF), is a single cylinder of titanium-doped sapphire about the width of a Frisbee. After kindling light in the crystal and shunting it through a system of lenses and mirrors, the SULF distills it into pulses of mind-boggling power. In 2016, it achieved an unprecedented 5.3 million billion watts, or petawatts (PW). The lights in Shanghai do not dim each time the laser fires, however. Although the pulses are extra-ordinarily powerful, they are also infinitesimally brief, lasting less than a trillionth of a second. The researchers are now upgrading their laser and hope to beat their own record by the end of this year with a 10-PW shot, which would pack more than 1000 times the power of all the world's electrical grids combined.

The group's ambitions don't end there. This year, Li and colleagues intend to start building a 100-PW laser known as the Station of Extreme Light (SEL). By 2023, it could be flinging pulses into a chamber 20 meters underground, subjecting targets to extremes of temperature and pressure not normally found on Earth, a boon to astrophysicists and materials scientists alike. The laser could also power demonstrations of a new way to accelerate particles for use in medicine and high-energy physics. But most alluring, Li says, would be showing that light could tear electrons and their antimatter counterparts, positrons, from empty space—a phenomenon known as "breaking the vacuum." It would be a striking illustration that matter and energy are interchangeable, as Albert Einstein's famous $E=mc^2$ equation states. Although nuclear weapons attest to the conversion of matter into immense amounts of heat and light, doing the reverse is not so easy. But Li says the SEL is up to the task. "That would be very exciting," he says. "It would mean you could generate something from nothing."

The Chinese group is "definitely leading the way" to 100 PW, says Philip Bucksbaum, an atomic physicist at Stanford University in Palo Alto, California. But there is plenty of competition. In the next few years, 10-PW devices should switch on in Romania and the Czech Republic as part of Europe's Extreme Light Infrastructure, although the project recently put off its goal of building a 100-PW-scale device (*Science*, 24 February 2017, p. 785). Physicists in Russia have drawn up a design for a 180-PW laser known as the Exawatt Center for Extreme Light Studies (XCELS), while Japanese researchers have put forward proposals for a 30-PW device.

Largely missing from the fray are U.S. scientists, who have fallen behind in the race to high powers, according to a study published last month by a National Academies of Sciences, Engineering, and Medicine group that was chaired by Bucksbaum. The study calls on the Department of Energy to plan for at least one high-power laser facility, and that gives hope to researchers at the University of Rochester in New York, who are developing plans for a 75-PW laser, the Optical Parametric Amplifier Line (OPAL). It would take advantage of beamlines at OMEGA-EP, one of the country's most powerful lasers. "The [Academies] report is encouraging," says Jonathan Zuegel, who heads the OPAL.

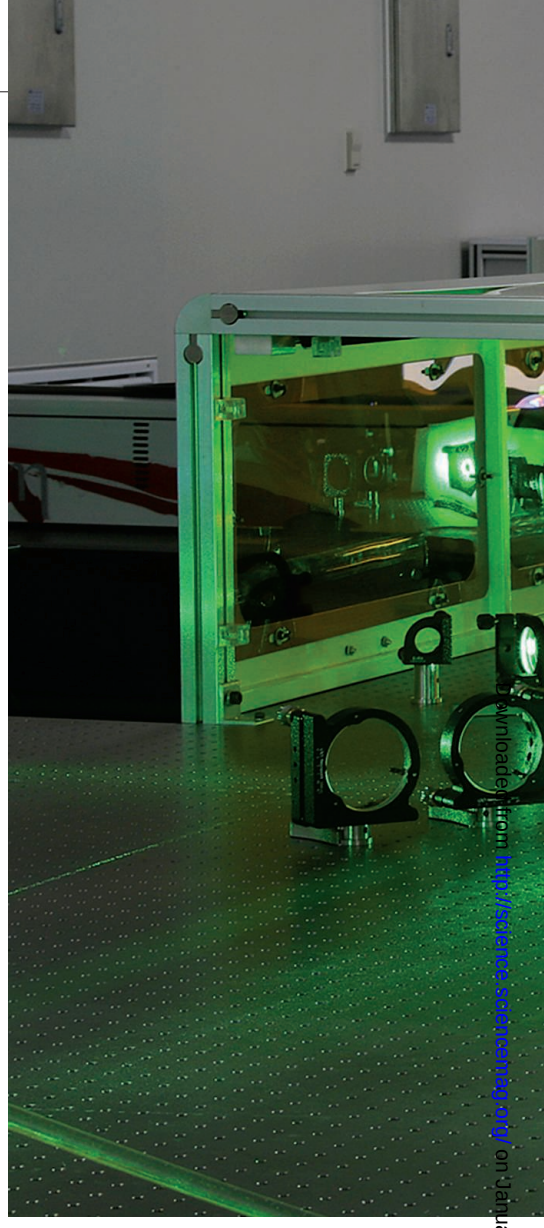
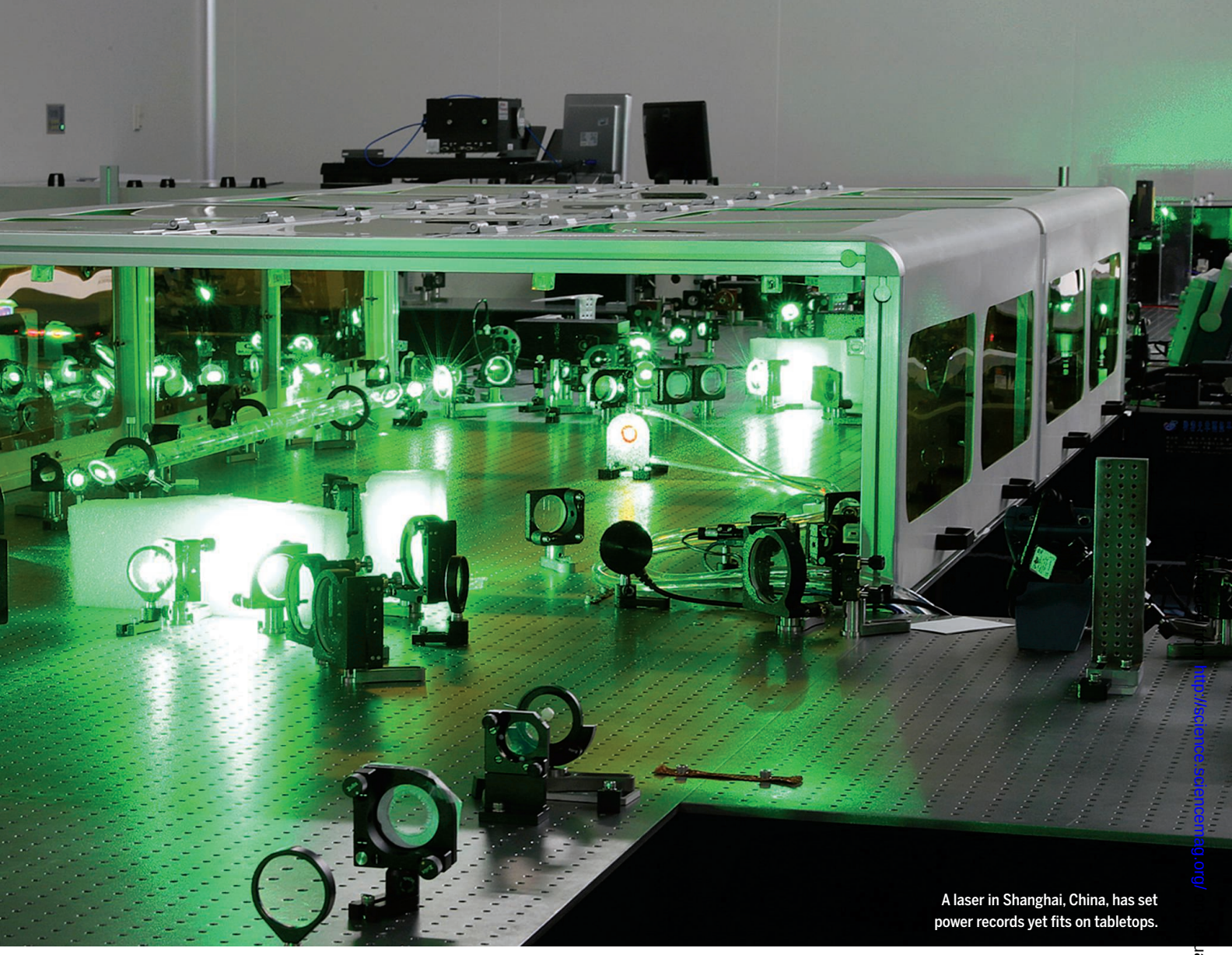


PHOTO: KAN ZHAN

INVENTED IN 1960, lasers use an external "pump," such as a flash lamp, to excite electrons within the atoms of a lasing material—usually a gas, crystal, or semiconductor. When one of these excited electrons falls back to its original state it emits a photon, which in turn stimulates another electron to emit a photon, and so on. Unlike the spreading beams of a flashlight, the photons in a laser emerge in a tightly packed stream at specific wavelengths.

Because power equals energy divided by time, there are basically two ways to maxi-



A laser in Shanghai, China, has set power records yet fits on tabletops.

mize it: Either boost the energy of your laser, or shorten the duration of its pulses. In the 1970s, researchers at Lawrence Livermore National Laboratory (LLNL) in California focused on the former, boosting laser energy by routing beams through additional lasing crystals made of glass doped with neodymium. Beams above a certain intensity, however, can damage the amplifiers. To avoid this, LLNL had to make the amplifiers ever larger, many tens of centimeters in diameter. But in 1983, Gerard Mourou, now at the École Polytechnique near Paris, and his colleagues made a breakthrough. He realized that a short laser pulse could be stretched in time—thereby making it less intense—by a diffraction grating that spreads the pulse into its component colors. After being safely amplified to higher energies, the light could be recompressed with a second grating. The end result: a more powerful pulse and an intact amplifier.

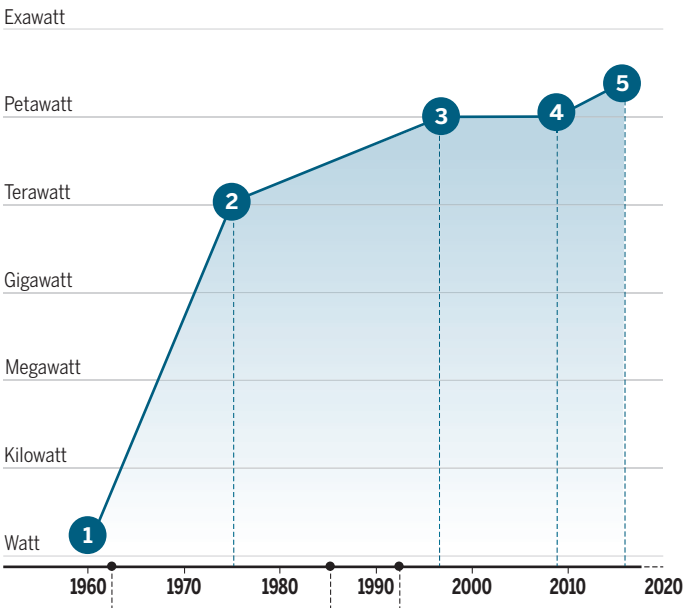
This "chirped-pulse amplification" has become a staple of high-power lasers. In 1996, it enabled LLNL researchers to generate the world's first petawatt pulse with the Nova laser. Since then, LLNL has pushed to higher energies in pursuit of laser-driven fusion. The lab's National Ignition Facility (NIF) creates pulses with a mammoth 1.8 megajoules of energy in an effort to heat tiny capsules of hydrogen to fusion temperatures. However, those pulses are comparatively long and they still generate only about 1 PW of power.

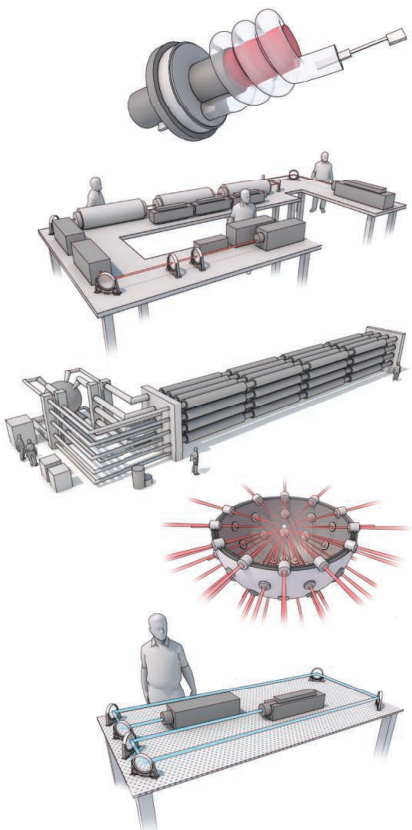
To get to higher powers, scientists have turned to the time domain: packing the energy of a pulse into ever-shorter durations. One approach is to amplify the light in titanium-doped sapphire crystals, which produce light with a large spread of frequencies. In a mirrored laser chamber, those pulses bounce back and forth, and the individual frequency components can be made to cancel each other out over most of their pulse length, while reinforcing each other in a fleeting pulse just a few tens of femtoseconds long. Pump those pulses with a few hundred joules of energy and you get 10 PW of peak power. That's how the SULF and other sapphire-based lasers can break power records with equipment that fits in a large room and costs just tens of millions of dollars, whereas NIF costs $3.5 billion and needs a building 10 stories high that covers the area of three U.S. football fields.

Raising pulse power by another order of magnitude, from 10 PW to 100 PW, will require more wizardry. One approach is to boost the energy of the pulse from hundreds to thousands of joules. But titanium-sapphire lasers struggle to achieve those energies because the big crystals needed for damage-free amplification tend to lase at right angles to the beam—thereby sapping energy from the pulses. So scientists at the SEL, XCELS, and OPAL are pinning their hopes on what are known as optical parametric amplifiers. These take a pulse stretched out by an optical grating and send it into an artificial "nonlinear" crystal, in

Powering up

Researchers at Lawrence Livermore National Laboratory (LLNL) in Livermore, California, set early power records by amplifying energies in mammoth machines. But a room-size laser in Shanghai, China, now holds the record, after squeezing modest energies into extremely short bursts. Three important techniques have propelled lasers to high powers.





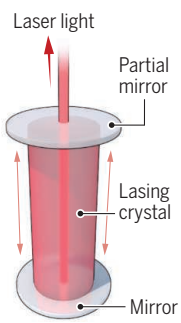
1 First laser
Theodore Maiman coaxed laser light from a 2-centimeter-long ruby crystal pumped by photographic flash lamps.

2 Janus (LLNL)
The two-beam laser amplified 100-picosecond pulses to 100 joules of energy to create the first terawatt shot.

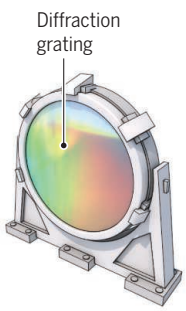
3 Nova (LLNL)
Pulses from the Nova laser were shortened using CPA to achieve the first petawatt.

4 National Ignition Facility (LLNL)
Shots focus 192 high-energy pulses on a target to induce fusion. Because the pulses are long, their power does not exceed a petawatt.

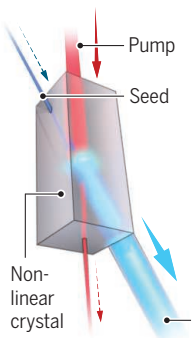
5 Shanghai Superintense Ultrafast Laser Facility
By squeezing laser pulses to just tens of femtoseconds, the laboratory achieved record powers with tabletop systems.



Mode locking
Although very pure, laser light is emitted over a range of wavelengths, or modes, that resonate in cavities like guitar strings. These modes can be made to constructively interfere for an intense burst tens of femtoseconds long.



Chirped-pulse amplification (CPA)
Intense pulses can damage amplifiers. CPA avoids that by stretching a laser pulse with diffraction gratings. After safe amplification, the pulse is compressed.



Optical parametric amplification
A high-energy pump beam can amplify a stretched seed pulse within a nonlinear crystal that can be made large to withstand intense inputs.

GRAPHIC: C. BICKEL/*SCIENCE*

which the energy of a second, "pump" beam can be channeled into the pulse. Recompressing the resulting high-energy pulse raises its power.

To approach 100 PW, one option is to combine several such pulses—four 30-PW pulses in the case of the SEL and a dozen 15-PW pulses at the XCELS. But precisely overlapping pulses just tens of femtoseconds long will be "very, very difficult," says LLNL laser physicist Constantin Haefner. They could be thrown off course by even the smallest vibration or change in temperature, he argues. The OPAL, in contrast, will attempt to generate 75 PW using a single beam.

Mourou envisions a different route to 100 PW: adding a second round of pulse compression. He proposes using thin plastic films to broaden the spectrum of 10-PW laser pulses, then squeezing the pulses to as little as a couple of femtoseconds to boost their power to about 100 PW.

Once the laser builders summon the power, another challenge will loom: bringing the beams to a singularly tight focus. Many scientists care more about intensity—the power per unit area—than the total number of petawatts. Achieve a sharper focus, and the intensity goes up. If a 100-PW pulse can be focused to a spot measuring just 3 micrometers across, as Li is planning for the SEL, the intensity in that tiny area will be an astonishing 10^{24} watts per square centimeter (W/cm^2)—some 25 orders of magnitude, or 10 trillion trillion times, more intense than the sunlight striking Earth.

THOSE INTENSITIES WILL OPEN the possibility of breaking the vacuum. According to the theory of quantum electrodynamics (QED), which describes how electromagnetic fields interact with matter, the vacuum is not as empty as classical physics would have us believe. Over extremely short time scales, pairs of electrons and positrons, their antimatter counterparts, flicker into existence, born of quantum mechanical uncertainty. Because of their mutual attraction, they annihilate each another almost as soon as they form.

But a very intense laser could, in principle, separate the particles before they collide. Like any electromagnetic wave, a laser beam contains an electric field that whips back and forth. As the beam's intensity rises, so, too, does the strength of its electric field. At intensities around 10^{24} W/cm^2, the field

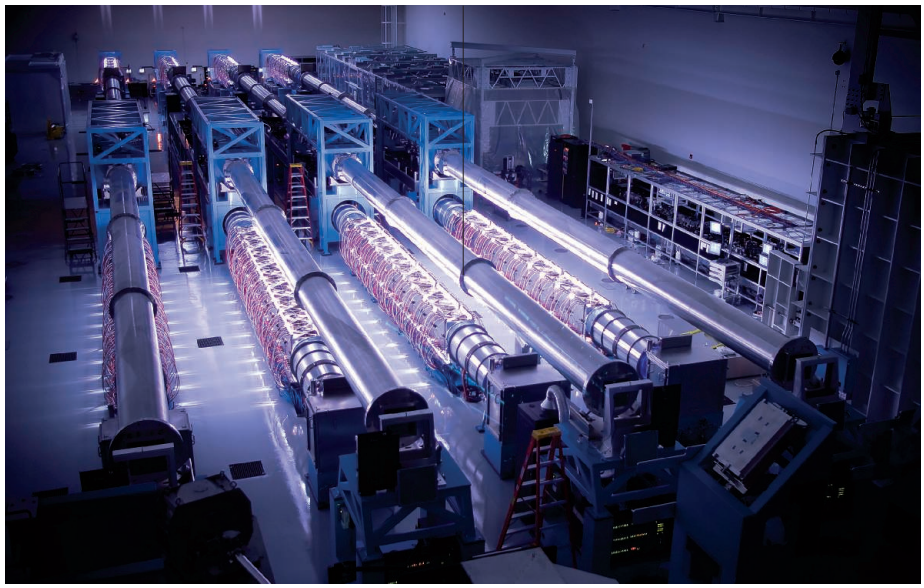
would be strong enough to start to break the mutual attraction between some of the electron-positron pairs, says Alexander Sergeev, former director of the Russian Academy of Sciences's (RAS's) Institute of Applied Physics (IAP) in Nizhny Novgorod and now president of RAS. The laser field would then shake the particles, causing them to emit electromagnetic waves—in this case, gamma rays. The gamma rays would, in turn, generate new electron-positron pairs, and so on, resulting in an avalanche of particles and radiation that could be detected. "This will be completely new physics," Sergeev says. He adds that the gamma ray photons would be energetic enough to push atomic nuclei into excited states, ushering in a new branch of physics known as "nuclear photonics"—the use of intense light to control nuclear processes.

One way to break the vacuum would be to simply focus a single laser beam onto an empty spot inside a vacuum chamber. But colliding two beams makes it easier, because this jacks up the momentum needed to generate the mass for electrons and positrons. The SEL would collide photons indirectly. First, the pulses would eject electrons from a helium gas target. Other photons from the laser beam would ricochet off the electrons and be boosted into high-energy gamma rays. Some of these in turn would collide with optical photons from the beam.

Documenting these head-on photon collisions would itself be a major scientific achievement. Whereas classical physics insists that two light beams will pass right through each other untouched, some of the earliest predictions of QED stipulate that converging photons occasionally scatter off one another. "The predictions go back to the early 1930s," says Tom Heinzl, a theoretical physicist at Plymouth University in the United Kingdom. "It would be good if we could confirm them experimentally."

BESIDES MAKING LASERS more powerful, researchers also want to make them shoot faster. The flash lamps that pump the initial energy into many lasers must be cooled for minutes or hours between shots, making it hard to carry out research that relies on plenty of data, such as investigating whether, very occasionally, photons transform into particles of the mysterious dark matter thought to make up much of the universe's mass. "Chances are you would need a lot of shots to see that," says Manuel Hegelich, a physicist at the University of Texas in Austin.

A higher repetition rate is also key to using a high-power laser to drive beams of particles. In one scheme, an intense beam would transform a metal target into a plasma, liberating electrons that, in turn, would eject protons from nuclei on the metal's surface. Doctors could use those proton pulses to destroy cancers—and a higher firing rate would make it easier to administer the treatment in small, individual doses.



Amplifiers for the University of Rochester's OMEGA-EP, lit up by flash lamps, could drive a U.S. high-power laser.

PHOTO: UNIVERSITY OF ROCHESTER LABORATORY FOR LASER ENERGETICS/EUGENE KOWALUK

Physicists, for their part, dream of particle accelerators powered by rapid-fire laser pulses. When an intense laser pulse strikes a plasma of electrons and positive ions, it shoves the lighter electrons forward, separating the charges and creating a secondary electric field that pulls the ions along behind the light like water in the wake of a speedboat. This "laser wakefield acceleration" can accelerate charged particles to high energies in the space of a millimeter or two, compared with many meters for conventional accelerators. Electrons thus accelerated could be wiggled by magnets to create a so-called free-electron laser (FEL), which generates exceptionally bright and brief flashes of x-rays that can illuminate short-lived chemical and biological phenomena. A laser-powered FEL could be far more compact and cheaper than those powered by conventional accelerators.

In the long term, electrons accelerated by high-repetition PW pulses could slash the cost of particle physicists' dream machine: a 30-kilometer-long electron-positron collider that would be a successor to the Large Hadron Collider at CERN, the European particle physics laboratory near Geneva, Switzerland. A device based on a 100-PW laser could be at least 10 times shorter and cheaper than the roughly $10 billion machine now envisaged, says Stuart Mangles, a plasma physicist at Imperial College London.

Both the linear collider and rapid-fire FELs would need thousands, if not millions, of shots per second, well beyond current technology. One possibility, being investigated by Mourou and colleagues, is to try to combine the output of thousands of quick-firing fiber amplifiers, which don't need to be pumped with flash tubes. Another option is to replace the flash tubes with diode lasers, which are expensive, but could get cheaper with mass production.

For the moment, however, Li's group in China and its U.S. and Russian counterparts are concentrating on power. Efim Khazanov, a laser physicist at IAP, says the XCELS could be up and running by about 2026—assuming the government agrees to the cost: roughly 12 billion rubles (about $200 million). The OPAL, meanwhile, would be a relative bargain at between $50 million and $100 million, Zuegel says.

But the first laser to rip open the vacuum is likely to be the SEL, in China. An international committee of scientists last July described the laser's conceptual design as "unambiguous and convincing," and Li hopes to get government approval for funding—about $100 million—early this year. Li says other countries need not feel left in the shadows as the world's most powerful laser turns on, because the SEL will operate as an international user facility. Zuegel says he doesn't "like being second," but acknowledges that the Chinese group is in a strong position. "China has plenty of bucks," he says. "And it has a lot of really smart people. It is still catching up on a lot of the technology, but it's catching up fast." ■

Edwin Cartlidge is a journalist in Rome.

PERSPECTIVES

GENETICS

Using nature to understand nurture

Genetic associations show how parenting matters for children's education

By **Philipp D. Koellinger**[1]
and **K. Paige Harden**[2]

Children resemble their parents in health, wealth, and well-being. Is parent-child similarity in traits and behaviors due to nature (the genes that children inherit from their parents) or nurture (the environment that parents provide for their children)? Answering this enduring question can directly inform our efforts to reduce social inequality and disease burden. On page 424 of this issue, Kong *et al.* (*1*) use genetic data from trios of parents and offspring to address this question in an intriguing way. By measuring parents' and children's genes, they provide evidence that inherited family environments influence children's educational success, a phenomenon termed genetic nurture.

Specifically, Kong *et al.* show that the part of the parental genotype that children

[1]*De Boelelaan 1085, 1081 HV Amsterdam, Netherlands.*
[2]*108 E. Dean Keeton Stop #A8000, Austin, TX 78704, USA.*
Email: p.d.koellinger@vu.nl; harden@utexas.edu

PHOTO: HERO IMAGES/GETTY IMAGES

Genetic nurture considers the influence of family environment on offspring.

do not inherit can nonetheless predict children's educational attainment. This genetic nurture effect is an indirect link between parental genotypes and children's characteristics, not caused by the children's own biology but rather by the family environment that covaries with parental genes. The concept of genetic nurture in human families extends previous work by behavioral ecologists and animal breeders on indirect genetic effects: environmentally mediated effects of one organism's genes on the phenotype of a related organism (*2*). For instance, one source of genetic variation in dung beetle body size is how large a brood mass the mother beetle produces (*3*). Interestingly, given how much of the animal research on indirect genetic effects has centered on offspring body size, Kong *et al.* found no evidence of genetic nurture on children's height and body mass index (BMI). For humans living in wealthy societies, genetic nurture might be more relevant for behavior and social achievements than for more biologically proximal outcomes such as body size.

Although the analyses of Kong *et al.* focus on a single child within a nuclear family, they are careful to remind us that family is more than one generation deep and often includes more than one child per generation. The environment that parents provide for their children could reflect the long arm of nurture by previous ancestors. And, siblings might create environments for each other. Genetic nurture could operate through any physical or social environment woven by genetic kin—a tangled web indeed.

The phenomenon of genetic nurture underscores that the results of genetic association studies cannot be interpreted as support for a biologically determinist account of human individual differences, as most social scientists who work with genetic data have stressed (*4*, *5*). Rather, results from genetic association studies are correlations that might point the way toward understanding causal mechanisms, but these mechanisms are likely to be complex and phenotype-specific. Genetic associations might depend on environmental context (*6*); for example, the alleles influencing educational attainment might differ between birth cohorts that have experienced different social and political institutions. And genetic effects can exert their influence via environmentally mediated channels, such as how teachers react to students with specific genotypes (*7*, *8*). Now, genetic nurture provides another compelling example of how tightly genetic and environmental mechanisms are entangled.

Genetic nurture also raises important methodological issues for future studies. Geneticists have long been concerned with population stratification (*9*), which is defined as the existence of systematic differences in gene frequency between groups that might also differ in environment and culture. Population stratification can induce a genotype-phenotype correlation that is, in fact, due to an unidentified group environment. The classic example is the chopsticks gene (*10*): Any genetic variant that differs in frequency between Asians and Europeans would come to be correlated with using chopsticks if population stratification were uncontrolled. Controlling for population stratification is typically accomplished by using statistical techniques that adjust for ancestry-specific genetic differences among people. However, there is no clear line between population stratification and genetic nurture, particularly once we realize that genetic nurture effects might themselves reproduce across generations: At what point does family history become ancestry? For researchers whose aim is to identify only those genetic effects that could be causally manipulated by changing an individual's own genotype, genetic nurture is, like population stratification, a source of false results.

"...genetic nurture provides another compelling example of how tightly genetic and environmental mechanisms are entangled."

Genetic nurture also presents a new challenge for research designs such as Mendelian randomization (*11*) that aim to identify causal effects by using genes as if they were naturally occurring experiments. For example, a researcher might use the alleles associated with educational attainment to test whether going further in school has a causal effect on one's health. One of the assumptions of Mendelian randomization is that genes are not correlated with unobserved confounds, such as parental environment. The study of Kong *et al.* clearly shows that this assumption can be wrong. If alleles identified in genetic association studies are not, in fact, independent of the parental environment, then Mendelian randomization studies might detect causal effects that are not really there.

Yet, genetic nurture does not undermine the value of genetic associations for prediction purposes. For example, researchers who want to control for genetic effects in order to obtain more precise estimates of an environmental or medical treatment in a randomized controlled trial (*12*) do not care why the genes are correlated with the outcome. All that matters is that the genes used as control variables predict as much variance as possible in the target outcome.

Furthermore, for many scientists nurture, not nature, is the phenomenon of interest. The study from Kong *et al.* provides these scientists with a fascinating new tool for investigating the effects of the family environment. Until now, the primary tools to disentangle the effects of a parent's genes from their actual parenting were adoption studies (*13*) and children-of-twins studies (*14*). However, collecting such data is difficult, and these study designs rarely represent the entire range of environments. Kong *et al.* capitalize on the same logic as an adoption study. The nontransmitted parental alleles function like the genotype of an adoptive parent, in that they help to shape the rearing environment but are independent from the offspring's genotype. Data sets with genotyped trios and high-quality measurements of the family environment are, unfortunately, still rare. But because of low genotyping costs, the trio design developed by Kong *et al.* could become a valuable and cost-effective research paradigm for scientists interested in understanding the impact of family environments on human flourishing.

The ingenious analysis of genetic trio data by Kong *et al.* reminds us yet again of the methodological problems that plague scientists as we try to understand individual differences in complex human behaviors and diseases—but also illustrates how understanding nature can provide us with new tools for studying nurture. ■

REFERENCES

1. A. Kong *et al.*, *Science* **359**, 424 (2018).
2. J. B. Wolf *et al.*, *Trends Ecol. Evol.* **13**, 64 (1998).
3. J. Hunt, L. W. Simmons, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6828 (2002).
4. C. A. Rietveld *et al.*, *Science* **340**, 1467 (2013).
5. A. Okbay *et al.*, *Nature* **533**, 539 (2016).
6. R. de Vlaming *et al.*, *PLOS Genet.* **13**, e1006495 (2017).
7. E. M. Tucker-Drob *et al.*, *Curr. Dir. Psychol. Sci.* **22**, 349 (2013).
8. C. Jencks, *Am. Sociol. Rev.* **45**, 723 (1980).
9. A. L. Price *et al.*, *Nat. Genet.* **38**, 904 (2006).
10. D. H. Hamer, *Mol. Psychiatry* **5**, 11 (2000).
11. D. A. Lawlor *et al.*, *Stat. Med.* **27**, 1133 (2008).
12. D. J. Benjamin *et al.*, *Annu. Rev. Econom.* **4**, 627 (2012).
13. J. M. Horn *et al.*, *Behav. Genet.* **9**, 177 (1979).
14. B. M. D'Onofrio *et al.*, *J. Child Psychol. Psychiatry Allied Discip.* **44**, 1130 (2003).

10.1126/science.aar6429

SOLAR CELLS

Perovskite solar cells must come of age

Developing aging standards is required for industrialization

By **Michael Saliba**

Perovskite solar cells (PSCs) have reached peak performances rivaling those of established technologies that have been painstakingly optimized for decades (*1*–*3*). Their high power outputs and low production costs have attracted serious industry attention from established companies and have led to the founding of multiple start-up companies (*4*). However, for commercial products, long-term stability is crucial. Thus, for perovskites to succeed, an informed discussion on a standard for stability data is required.

Characterizing perovskites is notoriously difficult because of the "hysteresis phenomenon" caused by mobile ions (see the figure, top panel) that may lead to the discrepancy between scans of the current density J for forward and reverse sweeps of voltage V (*5*, *6*). As shown in the bottom left figure panel, traditional fast JV scans (>100 mV/s) may overestimate power outputs. Slow scan speeds of ~10 mV/s are needed, together with maximum power point (MPP) tracking, where the solar cell is operated under full illumination and load-resembling working conditions (not illustrated in the figure).

Before the hysteresis phenomenon was fully appreciated, JV curves were frequently taken naïvely employing fast scan rates or preconditioning partially used for other materials (e.g., amorphous silicon), such as light soaking or prebiasing, altering ion migration substantially. From these measurements, "efficiency" values were extracted that were incomparable among research groups because of incomplete reporting of the precise testing conditions. The ambiguity in measuring efficiency, one of the most fundamental parameters, stifled the development of the research field. This is also the reason why certified perovskite efficiencies are still being classified as "not stabilized" (*1*).

With the JV question being slowly settled, there is an evident parallel with respect to stability measurements of perovskites. For example, using elevated temperature and light cycling for aging are well known for silicon; these tests are frequently conducted under nonoperational conditions (e.g., in the dark), because for silicon solar cells, such measurements have been established as a sufficient proxy for long-term stability (*7*). However, PSCs do not have established metrics for long-term stability extrapolation because perovskites are unusual semiconductors exhibiting ion migration that can take many hours to reach stable, steady-state conditions. Indeed, accelerated aging protocols for PSCs need to be developed as a function of standard aging parameters. Until then, all aging tests must be recorded and justified meticulously. Previous aging routines (e.g., for organic or silicon photovoltaics) serve better as guidelines than as a mandatory standard.

The suggested aging routine for PSCs in the figure (bottom right panel) acknowledges the nature of perovskite semiconductors. Because of the effects of ion migration, obtaining meaningful long-term stability data for PSCs requires testing under real working conditions. This entails an emphasis on using MPP tracking, where the solar cell is operated under illumination and load (*3*, *5*). Ion migration may cause an initial or "reversible loss" component. Thus, any long-term data require at least 100 hours, ideally even more, to distinguish the effects of short-term ion migration and long-term degradation. In this sense, "hot" or "cold" drawer tests—i.e., dark storage at ambient or elevated temperature, where the solar cell is not generating power—are unsuited to analyzing short-term versus long-term degradation (*8*).

Importantly, PSCs may recover after an overnight resting period without load and

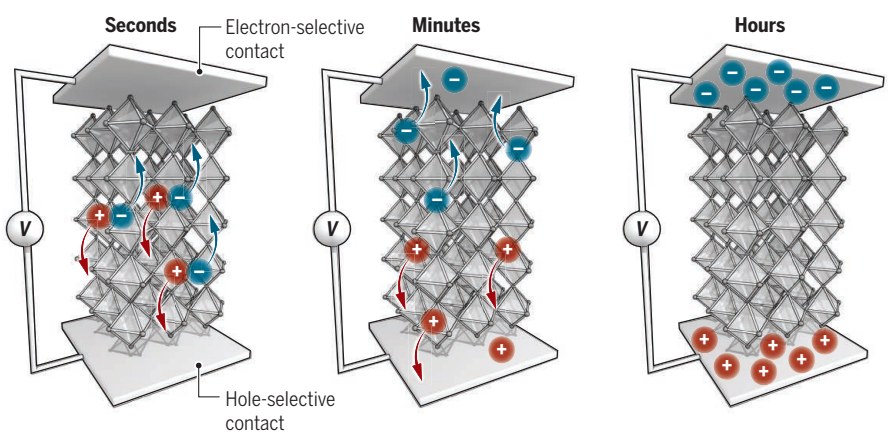
Adolphe Merkle Institute, University of Fribourg, CH-1700 Fribourg, Switzerland. Email: michael.saliba@unifr.ch

New aging standards for perovskite solar cells

Mobile ions cause hysteresis effects in measurements of these solar cells that can lead to artifacts in current density–voltage (JV) scans. Proposed stability measurements require full temperature and atmosphere control as well as maximum power point (MPP) tracking under constant illumination.

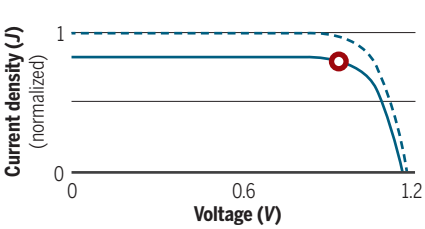
Mobile ions

Perovskites contain mobile ions that migrate toward the charge-selective contacts requiring hours to be in stable, steady-state working conditions.



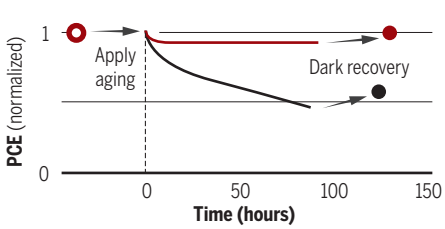
Scanning speeds

A fast JV scan is inaccurate (dashed blue line). A slow scan (solid blue line) starts resembling steady-state conditions. Realistic operation conditions are achieved with additional MPP tracking (open circle).



Aging and power conversion efficiency (PCE)

PCE of multiple devices (red and black lines) is recorded for at least 100 hours at MPP. All device parameters are measured before the aging routine is applied (open circle) and after aging with dark recovery (solid circles).



GRAPHIC: C. BICKEL/*SCIENCE*

illumination. Therefore, measurements should be taken before aging (open circle in bottom right figure panel) and after aging, including a resting period in the dark (solid circles in the bottom right figure panel) to determine the "reversible loss." Devices with different starting performances vary largely: A low-efficiency device may be considerably less stable than a high-efficiency counterpart (see red and black lines in the bottom right figure panel). Hence, PSC aging routines need to disclose all data with absolute rather than relative (or "normalized") values and ideally provide repeated measurements to ensure the validity of the data. Bare perovskite layers are very sensitive to extrinsic degradation, e.g., ambient moisture or oxygen (*9*). These can be remedied with encapsulation or, alternatively, an inert nitrogen atmosphere resembling "perfect encapsulation" (*10*, *11*), but these procedures and conditions must be stated as well. The outlined aging routine is only a starting point and may well change as the understanding of aging mechanisms in PSCs is improved. However, it is along these parameters that a calibration curve for accelerated aging can be developed to start the process of establishing a standard.

The power outputs from PSCs are sufficiently high to make them serious contenders for sustainable energy sources. In the coming years, PSCs will undergo relentless scrutiny for one of the most important topics in terms of market viability: long-term stability. Unfortunately, characterizing PSCs is challenging because of ion mobility. These issues already cast serious doubt on the reliability of efficiency measurements from the early period of perovskite research. A critical self-reflection on aging standards is needed to establish a basis for norms and common procedures with the chief aim of providing a calibration curve that fully captures the long-term stability of PSCs. Aging routines that account for the hour-long settling time of mobile ions, as outlined here, can then allow for the development of accelerated aging routines to guide industrial development. ■

REFERENCES

1. National Renewable Energy Laboratory, Best research-cell efficiencies; www.nrel.gov/pv/assets/images/efficiency-chart.png.
2. S. S. Shin *et al.*, *Science* **356**, 167 (2017).
3. M. Saliba *et al.*, *Science* **354**, 206 (2016).
4. M. A. Green, A. Ho-Baillie, *ACS Energy Lett.* **2**, 822 (2017).
5. H. J. Snaith *et al.*, *J. Phys. Chem. Lett.* **5**, 1511 (2014).
6. E. L. Unger *et al.*, *Energy Environ. Sci.* **7**, 3690 (2014).
7. Y. G. Rong, L. F. Liu, A. Y. Mei, X. Li, H. W. Han, *Adv. Energy Mater.* **5**, 1501066 (2015).
8. K. Domanski *et al.*, *Energy Environ. Sci.* **10**, 604 (2017).
9. J. W. Lee *et al.*, *Adv. Energy Mater.* **5**, 1501310 (2015).
10. F. Matteocci *et al.*, *Nano Energy* **30**, 162 (2016).
11. K. A. Bush *et al.*, *Adv. Mater.* **28**, 3937 (2016).

10.1126/science.aar5684

ANTHROPOLOGY

When did modern humans leave Africa?

A ~180,000-year-old fossil from Israel provides evidence for early forays of *Homo sapiens* into western Asia

By **Chris Stringer** *and* **Julia Galway-Witham**

The skeletal features of our species, *Homo sapiens,* include a globular braincase, brow ridges that are divided into central and side portions, a flat and retracted midface, a chin on the lower jaw, and a narrow pelvis. Fossils showing many of these characteristics have been excavated from the Ethiopian sites of Omo Kibish and Herto, dated at ~195,000 and ~160,000 years ago, respectively (*1*). Possible more primitive members of the species are known from Jebel Irhoud (Morocco) and Florisbad (South Africa), dated at ~315,000 and ~259,000 years ago, respectively (*1*). Yet, the oldest known *H. sapiens* fossils outside of Africa, from Skhul and Qafzeh in Israel, have been dated to just 90,000 to 120,000 years old. On page 456 of this issue, Hershkovitz *et al.* (*2*) provide fossil evidence from Misliya Cave, Israel, suggesting that our species had already left Africa by ~180,000 years ago (see the figure).

Misliya Cave lies on Mount Carmel, Israel, and has been excavated over the last century. It shows successive periods of hominin occupation with Lower and Middle Paleolithic tools, the use of fire, and the exploitation of a diverse fauna. Given Misliya's position at a crucial crossroads between Africa and Eurasia, the latest discoveries have enormous potential for understanding early populations of *H. sapiens* in western Asia.

The newly excavated material is represented by a partial upper jaw (Misliya-1), which includes some of the bone surrounding the tooth sockets, part of the cheekbone, the roof of the mouth, the bottom of the nasal cavity, and the complete upper left dentition. The size and shape of the specimen fall within the known range of variation of later *H. sapiens* fossils. Furthermore, Misliya-1 lacks the unique diagnostic features of Neandertals and earlier hominin species.

Hershkovitz *et al.* note Misliya-1's similarities to the later Skhul and Qafzeh fossils, although the size of one of the tooth cusps in Misliya-1 is reduced relative to the other fossils. Individual traits associated with the morphology exhibited in Misliya-1 are sometimes seen in other hominin taxa, but the combination of traits is characteristically that of *H. sapiens*.

Stone tools excavated from the same stratigraphic layer as that of Misliya-1 show the use of Levallois technology, a complex tool preparation method involving a prepared core. This technology has also been identified in ~190,000- to 260,000-year-old artifacts from nearby Tabun Cave (*3*), but the material at Misliya represents the earliest known association of this industry with modern human fossils in the region. Levallois tools associated with putative early *H. sapiens* fossils have also been found at Jebel Irhoud (Morocco) (*1*), suggesting that the emergence of this tool technology may be linked with the appearance and dispersal of our species in both Africa and western Asia.

> ***"...the latest discoveries have enormous potential for understanding early populations of* H. sapiens *in western Asia."***

Hershkovitz *et al.* dated the fossil and archaeological specimens from Misliya using multiple methods. They used uranium-thorium dating to determine the age of tooth dentine samples; combined uranium series and electron spin resonance techniques to directly date enamel samples; uranium-thorium dating to determine the age of the sediment adhering to the upper jaw; and thermoluminescence to measure the age of the burned tools found close to the fossil. Collectively, the results of these dating methods provide an estimated age of ~177,000 to 194,000 years for Misliya-1 and its associated artifacts.

It remains unclear where and when humans lived in western Asia in the period prior to that of the Misliya specimen. Fragmentary

Department of Earth Sciences, Natural History Museum, London SW7 5BD, UK. Email: c.stringer@nhm.ac.uk; j.galway-witham@nhm.ac.uk

fossils are known from the Israeli caves of Qesem and Zuttiyeh, the former consisting of isolated teeth dated to ~400,000 years old, the latter represented by a partial skull, perhaps of similar antiquity. Both show primitive traits, and some that are found in Neandertals or *H. sapiens* (*4, 5*). Archaeological evidence for occupation of the region spans the time between these fossils and Misliya, but imprecise dating makes it impossible to tell whether human presence was continuous or episodic.

Paleoclimatic reconstructions using speleothems, deep sea cores, and paleoenvironmental data suggest that there were several humid phases between 244,000 and 190,000 years ago, one or more of which could have facilitated the spread of *H. sapiens* into the region (*6*). But there were severe periods of aridity before and after this time, meaning that the region was probably more often a "boulevard of broken dreams" than a stable haven for early humans (*7*). Direct local evolutionary continuity between the population represented by the Misliya fossil and later Skhul and Qafzeh peoples thus seems unlikely.

If the Misliya population extended further into Eurasia, encountering Neandertals, this may have led to gene flow between these two lineages of humans. The main phases of genetic introgression from Neandertals into *H. sapiens* are estimated to have occurred between ~50,000 and 60,000 years ago (*8*), but genetic analyses of Neandertal fossils from Denisova Cave (Siberia, Russia) and Hohlenstein-Stadel (Germany) indicate at least one earlier phase of introgression, from *H. sapiens* into Neandertals. This event has been estimated at ~219,000 to 460,000 years ago, suggesting that it predated the Misliya fossil (*9*). The *H. sapiens* fossils from Misliya, Skhul, and Qafzeh could therefore represent relatively late excursions of our species from Africa. During earlier humid phases, western Asia could have attracted more primitive members of the *H. sapiens* and *H. neanderthalensis* lineages, providing repeated opportunities for exchanges of genes and technologies.

Beyond western Asia, evidence for early dispersals of *H. sapiens* prior to ~120,000 years ago is weak. Age estimates for fragmentary Chinese fossils from Zhiren and Daoxian Caves only range between ~80,000 and 113,000 years ago (*10*). The Misliya find is important in establishing an earlier dispersal of *H. sapiens* from Africa into western Asia. It also highlights how little we know about the early occupants of the region, nearly a century after the first discoveries from the caves of Zuttiyeh, Skhul, Qafzeh, and Tabun. ■

REFERENCES AND NOTES

1. C. B. Stringer, J. Galway-Witham, *Nature* **546**, 212 (2017).
2. I. Hershkovitz *et al.*, *Science* **359**, 456 (2018).
3. N. Mercier, H. Valladas, *J. Hum. Evol.* **45**, 401 (2003).
4. I. Hershkovitz *et al.*, *Am. J. Phys. Anthropol.* **144**, 575 (2011).
5. S. E. Freidline *et al.*, *J. Hum. Evol.* **62**, 225 (2012).
6. P. S. Breeze *et al.*, *Quat. Sci. Rev.* **144**, 155 (2016).
7. J. J. Shea, in *Rethinking the Human Revolution*, P. Mellars *et al.*, Eds. (McDonald Institute for Archaeological Research Monographs, Cambridge, 2007), chap. 19.
8. R. Nielsen *et al.*, *Nature* **541**, 302 (2017).
9. C. Posth *et al.*, *Nat. Commun.* **8**, 16046 (2017).
10. M. Martinón-Torres *et al.*, *Curr. Anthropol.* **58**, S434 (2017).

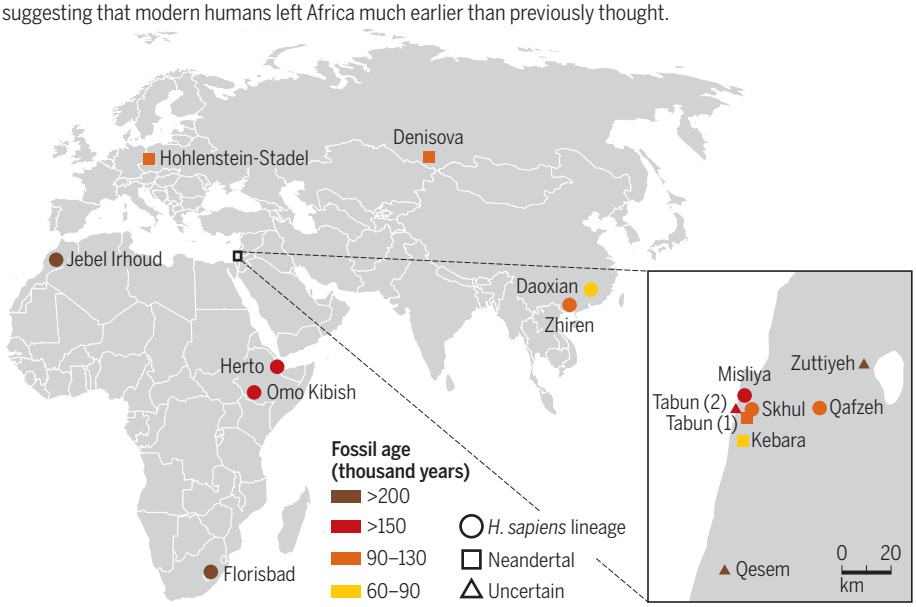
ACKNOWLEDGMENTS

Authors are supported by the Calleva Foundation and the Human Origins Research Fund.

10.1126/science.aas8954

Earliest modern human migrations from Africa

Hershkovitz *et al*. report an age of ~180,000 years for a modern human fossil from Misliya in Israel, suggesting that modern humans left Africa much earlier than previously thought.



GRAPHIC: J. YOU/*SCIENCE*

CELL BIOLOGY

Complexity in targeting membrane proteins

Discovery of a new pathway provides a role for a conserved membrane protein complex

By **Michelle Y. Fry** *and* **William M. Clemons Jr.**

For a cell, making hydrophobic integral membrane proteins (IMPs, or transmembrane proteins) is a complicated, but critical, process. All proteins are made by ribosomes in the cytosol, but IMPs, which account for ~30% of the proteins encoded in the eukaryotic genome, must also be properly delivered to and inserted into their respective subcellular membranes, a process known as targeting. This is important because hydrophobic IMPs are rapidly degraded in the cytoplasm to prevent aggregation, which can lead to broad disruptions in cellular homeostasis. Because of the number and diversity of IMPs, identifying and targeting them relies on pathways that often overlap in function. The information for targeting, typically stored in hydrophobic α-helical transmembrane domain (TMD) signals in the IMPs, is recognized by factors that then ferry the IMP client to the destined lipid bilayer. On page 470 of this issue, Guna *et al*. (*1*) demonstrate that, for an essential class of IMPs, the current picture is incomplete. They identify a new targeting pathway for these IMPs, contributing to the elucidation of how the cell synthesizes this important class of proteins.

IMP targeting is dominated by the secretory (SEC) pathway, with the signal-recognition particle (SRP) being the central targeting factor. For most IMPs, the SRP binds the N-terminal signal as it emerges from a ribosome. The subsequent nascent protein–SRP complex is delivered to the endoplasmic reticulum (ER) membrane for cotranslational insertion via the SEC translocon, an ER-

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA. Email: clemons@caltech.edu

protein complex that acts as a conduit for the insertion of IMPs into the ER membrane (*2*). However, not all IMPs can access the SEC pathway. The tail-anchored (TA) IMPs, characterized by a single TMD at their carboxyl terminus, are a large class of SEC-independent IMPs that are initially targeted to either the ER or mitochondrial membranes. The TMD is the signal that targets these TA IMPs for membrane insertion, yet it remains hidden within the ribosomal exit tunnel at the end of translation. This necessitates posttranslational targeting of the newly synthesized protein. The first pathway identified specifically for TA IMP targeting, dubbed the TMD recognition complex (TRC) pathway in mammals [or the guided entry of TA proteins (GET) pathway in yeast], delivers TA clients to the ER with the adenosine triphosphatase TRC40 (Get3 in yeast) as the central targeting factor (*2*–*4*).

In the cytoplasm, any TMDs that emerge from ribosomes are quickly bound by cytoplasmic factors. For TA IMPs targeted by the TRC pathway, the co-chaperone SGTA (small glutamine-rich tetratricopeptide repeat-containing alpha) binds TMDs before hand-off to TRC40 (*5*). Another TMD binder is the conserved calcium-sensing protein calmodulin, which also binds to small secreted proteins and short signal peptides (*2, 6*). Both SGTA and calmodulin, like the SRP, have a methionine-rich domain for recognition of a range of hydrophobic substrates. TRC40 can capture TMDs of TA IMPs released from calmodulin when no other chaperones are present, but in physiological conditions, SGTA binds the TMD and directly hands off the IMP to TRC40 (*6*) (see the figure).

Deletion of the TRC pathway components in yeast results in phenotypes that include defects in Golgi-to-ER targeting, sensitivity to metal ions, and effects on the protein degradation machinery, which have been attributed to the mislocalization of TA IMPs (*3, 4*). When characterized in detail, deletion of TRC pathway components had some unexpected twists. In yeast, the localization of client TA IMPs varies in sensitivity to loss of GET proteins; some TA IMPs, such as Sed5, remain in the cytosol, whereas others, such as Sec61 and Emp47, localize correctly in membranes (*4, 7*). Consequently, there must be an alternative pathway(s) to traffic some TA IMPs. The recently discovered SRP-independent (SND) pathway has the ability to traffic several IMPs to the ER membrane, although the SND pathway may play a secondary role for some (*8*). The TMDs of TA IMPs vary in length and hydrophobicity, which likely explains the dependence on different pathways. How the various pathways discriminate between substrates remains an open question.

The GET pathway was identified in yeast by a series of beautiful genetic experiments which also found a group of proteins that form the ER membrane complex (EMC) (*9, 10*). This was intriguing because this large complex—comprising 10 subunits (EMC1 to 10) in mammals—is conserved in the earliest eukaryotic ancestors (*11*). The biochemical role of the EMC is unknown. Upon its discovery, Jonikas *et al.* (*10*) noted that loss of the EMC resulted in the accumulation of misfolded IMPs. Subsequently, the EMC has been implicated in multiple cellular functions, including the biogenesis of IMPs, exchange of lipids between the ER and mitochondria, ER-associated degradation, and tethering between the ER and mitochondria (*12*–*14*). Uncovering a specific role for the EMC will shed light on whether these functions are directly linked to loss of the EMC.

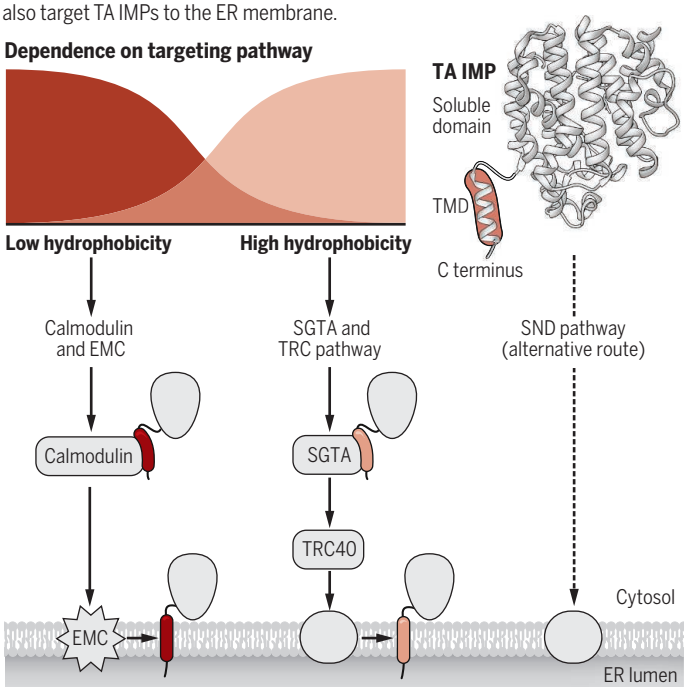
Guna *et al.* resolve, at least partially, some of these mysteries. They demonstrate that the TRC pathway fails to target ER-bound TA IMPs with low-hydrophobicity TMDs. Instead, this class of TA IMPs are clients for calmodulin, which is the only factor bound to such proteins after they are translated. Calmodulin–TA IMP complexes were sufficient to facilitate insertion into ER-derived membranes, but were dependent on the EMC. A fully reconstituted system in vitro demonstrated that purified EMC is necessary and sufficient for calmodulin-mediated TA IMP insertion into membranes. Thus, Guna *et al.* conclude that at least one prominent role for the EMC is to act as an IMP membrane insertase.

These results open up additional possibilities. Requirement of the EMC for targeting perhaps as much as 50% of TA IMPs suggests that the observed effects EMC deletion, such as the accumulation of misfolded IMPs, could indirectly result from failed TA IMP insertion, similar to defects arising from the deletion of the TRC or GET pathways. For example, a number of ER-bound TA IMPs are involved in lipid synthesis and ER-associated degradation. A direct role for the EMC in IMP membrane insertion is possible but remains to be demonstrated. Although other factors are not required, the lack of a direct interaction between calmodulin and the EMC suggests there may be accessory factors that remain to be determined.

The addition of this new pathway increases the level of complexity of IMP targeting. Although the findings of Guna *et al.* answer a number of questions, more protein-sorting pathways remain unknown; for example, we currently do not know how TA IMPs are targeted to mitochondria. Moreover, even for well-known pathways, new twists continue to be revealed. This is exemplified by Costa *et al.* (*15*), who demonstrate unexpected differences in dependencies of various IMPs on the SRP for correct targeting. Clearly, the future holds many more exciting discoveries in IMP targeting. ■

Fates of tail-anchored integral membrane proteins

TA IMPs that are destined to the ER membrane bind to calmodulin or SGTA, depending on their hydrophobicity. Calmodulin delivers low-hydrophobicity TA IMPs to the EMC for membrane insertion. More-hydrophobic TA IMPs are targeted through the SGTA-TRC pathway for membrane insertion. The SND pathway can also target TA IMPs to the ER membrane.



GRAPHIC: K. SUTLIFF/*SCIENCE*

REFERENCES

1. A. Guna *et al.*, *Science* **359**, 470 (2018).
2. S. Shao, R. S. Hegde, *Cell* **147**, 1576 (2011).
3. S. Stefanovic, R. S. Hegde, *Cell* **128**, 1147 (2007).
4. M. Schuldiner *et al.*, *Cell* **134**, 634 (2008).
5. F. Wang *et al.*, *Mol. Cell* **40**, 159 (2010).
6. S. Shao *et al.*, *Science* **355**, 298 (2017).
7. J. Rivera-Monroy *et al.*, *Sci. Rep.* **6**, 39464 (2016).
8. N. Aviram *et al.*, *Nature* **540**, 134 (2016).
9. M. Schuldiner *et al.*, *Cell* **123**, 507 (2005).
10. M. C. Jonikas *et al.*, *Science* **323**, 1693 (2009).
11. J. G. Wideman, *F1000Res* **4**, 624 (2015).
12. S. Lahiri *et al.*, *PLOS Biol.* **12**, e1001969 (2014).
13. R. J. Louie *et al.*, *Genome Med.* **4**, 103 (2012).
14. J. C. Christianson *et al.*, *Nat. Cell Biol.* **14**, 93 (2012).
15. E. A. Costa *et al.*, *Science* 10.1126/science.aar3607 (2018).

10.1126/science.aar5992

ECOLOGY

Conserving honey bees does not help wildlife

High densities of managed honey bees can harm populations of wild pollinators

By **Jonas Geldmann** *and* **Juan P. González-Varo**

There is widespread concern about the global decline in pollinators and the associated loss of pollination services. This concern is understandable given the importance of pollinators for global food security; ~75% of all globally important crops depend to some degree on pollination, and the additional yield due to pollination adds ~9% to the global crop production (*1*). These services are delivered by a plethora of species, including more than 20,000 species of bees as well as butterflies, flies, and many species of vertebrates (*1*). Yet, concern has focused on one species above all: the western honey bee (*Apis mellifera*). This is unfortunate because research shows that managed honey bees can harm wild pollinator species, providing an urgent incentive to change honey bee management practices.

The western honey bee is the most important single species for crop pollination, with a rapid global growth in managed colony numbers over the past decades, particularly in much of its introduced range. Honey production can also be an important source of income, particularly in many rural communities. Lack of pollination of commercial crops associated with the current honey bee die-off in some countries—most notably, the United States—is, however, an issue of agricultural rather than environmental importance.

Despite this, news stories often view honey bee losses through the lens of environmental concern (*2*). This has led to initiatives, masked as conservation, that promote honey bees in cities and even in protected areas far from agriculture (see the photo) (*3*). Nongovernmental organizations have even responded to the pollinator crisis with a call to action that includes suggestions to buy local honey and support honey bee conservation (*4*, *5*).

Across organizations and strategies, there is a recognition that there are pollinators other than the western honey bees. Nevertheless, the general belief that addressing the decline in managed honey bees would be an environmental feat persists in the media (*2*) and among the public (*6*). This lack of distinction between the declines of wild pollinators and the plight of a heavily managed, agricultural species may even reduce efforts to conserve wild pollinator species, many of which are nationally or even globally threatened (*1*).

Furthermore, there is increasing evidence that unnaturally high densities of honey bees, associated with beekeeping, can exacerbate declines in wild pollinators (*7*). This problem is particularly evident in areas where western honey bees have been introduced (*7*); but even in their native range in Europe, managed honey bees have been shown to depress the densities of wild pollinators around apiaries both in natural habitats (*8*) and in crop fields (*9*). Furthermore, they move toward surrounding natural habitats in unnaturally high densities after the blooming period of mass-flowering crops (*10*), potentially outcompeting wild pollinators (*11*).

Besides competing with wild species for resources (*12*), honey bees are linked to the spread of diseases to wild pollinators via shared flowers, an effect that is likely amplified by trade with and movement of honey bees (*13*). Honey bees can also have a negative impact on the reproductive success of wild plants (*11*) and even depress nonpollinator species—for example, the threatened Lear's Macaw in Brazil, which competes with honey bees for nest sites in rock cavities. The western honey bee thus unequivocally fits Geslin and colleagues' concept of a "massively introduced managed species," which, regardless of whether they are native or not, can negatively affect their environment through sheer numbers (*14*).

We do not dispute that managed honey bees are a useful, even a necessary agricultural tool for improving the yield of many mass-flowering crops. Additionally, many factors that negatively affect managed honey bees (such as neonicotinoids, parasites, and diseases) are also harming native pollinators. Thus, honey bees can serve as a "canary in a coalmine"; the declines observed in managed colonies across Europe and the United States are likely mirrored by many wild pollinator species. In fact, strategies developed to reduce managed honey bee losses, such as banning neonicotinoids, will also benefit many wild pollinators desperately in need of conservation attention.

But managed honey bees are a means, not an end, and strategies to ensure sufficient crop pollination need to take account of potential competition with native wild pollinators. This necessitates a better assessment of when, where, and in what densities honey bees are required to ensure effective pollination of mass-flowering crops without harming wild native pollinators or plants. Such assessments must

In the Teide National Park, Spain, up to 2700 honey bee hives are introduced each spring for honey production.

PHOTO: ALFREDO VALIDO

Conservation Science Group, Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK. Email: jg794@cam.ac.uk; jpg62@cam.ac.uk

explicitly account for the contributions of native wild pollinators, which may be responsible for as much as 50% of the needed pollination services (*15*). Within their native range, some amount of pollination by western honey bees is natural, although the historic density of wild colonies is largely unknown. Safe densities of managed honey bees will vary from natural and protected habitats, where wild native pollinators are most abundant and beekeeping is mainly done for honey production, to agricultural and managed landscapes, which are less important for the conservation of the most threatened pollinator species.

Management practices must also address the periods when no or insufficient mass-flowering crops are in bloom because managed honey bees are likely to compete most intensively with wild native pollinators during these times. In the United States, honey bee hives are moved around to track the bloom of various crops, from California almond groves in early spring to Washington apples in the late summer. Similar approaches might be needed across Europe and other places to match pollinator supply to pollination demand but must address the risk of spreading diseases. Policies to limit the number of honey bees in specific periods might also be needed, such as early honey removal and keeping the individual hives smaller. If implemented wisely, such strategies will come with no extra cost to farmers but may increase the price of honey.

Fulfilling the need for sufficient and effective pollination of the world's crops without jeopardizing biodiversity will also require an ambitious research agenda. The past decade has seen an explosion in research tackling the decline in managed honey bees, specifically focused on the potential loss of pollination services. This research has been heavily supported by the private sector and governments, particularly in Europe and the United States, which have invested millions to reverse the loss of managed honey bees. Comparatively little research has been undertaken to understand wild native pollinator declines, including the potential negative role of managed honey bees. The European project STEP (Status and Trends in European Pollinators; www.step-project.net), which aimed to document the nature and extent of pollinator declines and brought together 21 universities and institutions from 16 countries, exemplifies the type of research initiative needed to elucidate the drivers of pollinator declines.

Concern about honey bees has been an engine for shining light on the decline of pollinators and has likely been important in raising awareness of pollinator declines at large (*4*, *5*). Thus, a more nuanced understanding of the role of domesticated honey bees must not be misconstrued as a general lack of importance of conservation attention on wild native pollinators. Half of all European bees are threatened with extinction (*1*), and the conservation of wild native pollinators is among the most important conservation challenges in many parts of the world. We therefore see a need for a conservation strategy that explicitly focuses on the main drivers of the current declines in wild native pollinators, not on agricultural yield.

As a first step, crop pollination by managed honey bees should not be considered an ecosystem service because those pollination services are delivered by an agricultural animal and not by the local ecosystems. Further, managed honey bee hives should not be placed in protected areas, where they are likely to do the biggest damage to wild pollinators. In other areas of conservation importance, beekeeping may require impact assessments that consider potential spillover after the bloom of adjacent mass flowering crops. Honey bees may be necessary for crop pollination, but beekeeping is an agrarian activity that should not be confused with wildlife conservation. ■

REFERENCES AND NOTES

1. Intergovernmental Platform on Biodiversity and Ecosystem Services, *The Assessment Report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services on Pollinators, Pollination and Food Production* (Secretariat of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, 2016).
2. D. Bolton, Independent Digital News & Media, 2016; www.independent.co.uk/environment/bee-decline-dying-out-honeybees-uk-food-production-extinction-a6939266.html.
3. A. Valido, M. C. Rodriguez-Rodriguez, P. Jordano, *Ecosistemas* **23**, 58 (2014).
4. The Nature Conservancy, *Bees and Agriculture* (The Nature Conservancy, 2017); www.nature.org/ourinitiatives/regions/northamerica/unitedstates/indiana/journeywithnature/bees-agriculture.xml.
5. Greenpeace, SOS-bees (Greenpeace, 2014); http://sos-bees.org/situation.
6. YouGov, Decline of bees seen as more serious than climate change, YouGov, 2014; https://yougov.co.uk/news/2014/06/27/bees-dying-most-serious-environmental-issue.
7. R. E. Mallinger, H. R. Gaines-Day, C. Gratton, *PLOS ONE* **12**, e0189268 (2017).
8. A. Torné-Noguera, A. Rodrigo, S. Osorio, J. Bosch, *Basic Appl. Ecol.* **17**, 199 (2016).
9. S. A. M. Lindström, L. Herbertsson, M. Rundlöf, R. Bommarco, H. G. Smith, *Proc. R. Soc. B Biol. Sci.* **283**, 20161641 (2016).
10. J. P. González-Varo, M. Vilà, *Biol. Conserv.* **212**, 376 (2017).
11. A. Magrach, J. P. González-Varo, M. Boiffier, M. Vilà, I. Bartomeus, *Nat. Ecol. Evol.* **1**, 1299 (2017).
12. J. H. Cane, V. J. Tepedino, *Conserv. Lett.* **10**, 205 (2017).
13. M. A. Fürst, D. P. McMahon, J. L. Osborne, R. J. Paxton, M. J. F. Brown, *Nature* **506**, 364 (2014).
14. B. Geslin *et al.*, *Adv. Ecol. Res.* **57**, 147 (2017).
15. L. A. Garibaldi *et al.*, *Science* **339**, 1608 (2013).

ACKNOWLEDGMENTS

We thank L. V. Dicks and A. Valido for comments on the manuscript. J.G. (H2020-MSCA-IF-2015-706784) and J.P.G.-V. (H2020-MSCA-IF-2014-656572) were funded by Individual Marie Sklodowska-Curie Fellowships.

10.1126/science.aar2269

QUANTUM INFORMATION

Toward a silicon-based quantum computer

A controlled NOT gate for two quantum bits is demonstrated with a strained-silicon device

By **Lars R. Schreiber** *and* **Hendrik Bluhm**

Quantum computing could enable exponential speedups for certain classes of problems by exploiting superposition and entanglement in the manipulation of quantum bits (qubits). The leading quantum systems that can be used include trapped ions, superconducting qubits, and spins in semiconductors. The latter are considered particularly promising for scaling to very large numbers of qubits. On page 439 of this issue, Zajac *et al.* (*1*) demonstrate a quantum operation involving two qubits in silicon (Si), which is a major step for the field of semiconductor qubits. Together with easier-to-achieve manipulation of single qubits, these operations represent the basic steps of any quantum algorithm.

"...creating systems that cannot be simulated with today's supercomputers will take about 50 qubits."

The coupling between the two qubits is achieved through the so-called exchange interaction, which results from coupling of the two electrons through a tunnel barrier. This barrier can be controlled by changing the voltage on the central gate. The authors further use microwave excitation to implement the desired operation. In an external magnetic field, spins that are not aligned with the field precess around it like an

Jülich Aachen Research Alliance (JARA), Institute for Quantum Information, RWTH Aachen and Forschungszentrum Jülich, Germany. Email: lars.schreiber@physik.rwth-aachen.de

asymmetrically suspended top. If an excitation (a microwave signal applied to one of the gates, which translates to a magnetic microwave field in the inhomogeneous stray field of a micromagnet) has the same frequency as the precession, it is possible to rotate the spin direction, for example, from parallel to antiparallel with the field.

This technique, commonly used to control individual qubits, enables two-qubit operation via the dependence of the precession rate of one spin on the state of the other because of the exchange interaction. Whether single-qubit or two-qubit operation is executed can be selected by the choice of tunnel coupling and microwave frequency. For the two-qubit operation, they are set such that if the so-called control spin is aligned with the external field, the other so-called target spin is on resonance with the microwave signal (and off resonance if anti-aligned). Whether the target spin is inverted depends on the state of the control spin, so this system functions as a controlled NOT (CNOT) gate.

Zajac *et al.* used a device consisting of a layer of Si that was strained by being grown between two layers of SiGe (Si/SiGe), which confine electrons to the Si layer. Additional lateral confinement was provided by electrostatic gates fabricated on top of the structure, which were arranged such that two electrons can be captured (see the figure). The spin of each of those electrons encoded one qubit. Similar qubits have been realized in GaAs/AlGaAs heterostructures, but in that material system, the interaction with unavoidable nuclear spins is a major complication that impedes highly accurate qubit operation (*2*). Only 4.7% of the nuclei carry spin in Si with a natural composition, and that fraction can be further reduced with isotopic purification. This approach has recently led to record-setting coherence times over which quantum states could be preserved (*3*). However, the controlled confinement of single electrons in Si has been a major challenge because of disorder in the material. Petta and coworkers (*4*) made important progress on sample quality and design, culminating in an array of nine quantum dots, which could in principle host nine qubits.

A key figure of merit for any qubit is the gate fidelity, which specifies how accurately it can be manipulated. Values in excess of the 99.9% thought to be needed for quantum error corrections have now been shown for Si-based spin qubits in Si/SiGe (*5*) and in Si metal-oxide-semiconductor (MOS) structures, in which the electrons are confined at a Si/SiO_2 interface (*6*). Two-qubit operations required for quantum algorithms have already been demonstrated in Si MOS (*7*) but were not characterized as thoroughly as is now done by Zajac *et al.* Moreover, in parallel with the Vandersypen group (*8*), they have now demonstrated a two-qubit operation in Si/SiGe.

As a measure of the accuracy of the operation, the authors use a standard method (state tomography) to determine how well they can produce entangled states. The resulting fidelity (i.e., similarity with the desired state) of 78% is still far from the required 99.9%. Part of the reason is likely measurement uncertainties that can be reduced by using technically more involved, but well-established, techniques to determine the qubit state. Furthermore, it is likely that charge, and possibly nuclear spin noise, plays a role, given that the authors used a sample with natural isotopic composition. Understanding and substantially reducing these sources of errors, e.g., by optimized pulse sequences, are the next challenges to overcome. In GaAs, for example, optimized pulse sequences pushed the single-qubit gate fidelity near the predicted values (*9*). The ultimate limit will likely depend on charge noise, which is currently poorly understood and characterized.

In the near future, the results of Zajac *et al.* should be readily applicable to their nine-qubit array. For comparison, creating systems that cannot be simulated with today's supercomputers will take about 50 qubits. The most promising applications, such as solving quantum chemistry problems or the factorization of large numbers, are expected to require a redundant encoding of information to enable quantum error correction, which increases the number of qubits required to some millions. Thus, scalable quantum computing architectures with a fabrication yield as achieved in industrial silicon foundries are necessary.

It is encouraging that today's microprocessors have a comparable complexity. Nevertheless, the remaining challenges are substantial. For the quantum layer, long-range coupling of qubits at least a few micrometers apart will likely be required to make space for wiring or on-chip classical control electronics operating at cryogenic temperatures (*10*). Possible approaches include microwave cavities (*11*, *12*) and shuttling around electrons while preserving the quantum state of their spins. None of these have achieved the required fidelity yet. Finally, the creation of sufficiently powerful control systems will mark a whole new era of quantum computing research. ■

Manipulating electron spins

A schematic of the two-electron–trap sample used by Zajac *et al.* is shown. Negatively biased electrostatic gates (gray) form tunnel barriers, whereas positively biased gates (brown) accumulate electron reservoirs or exactly two electrons in the double-dot potential. Each electron spin qubit can be separately set in quantum superposition state, or the two qubits can be entangled with the microwave and voltage pulses.

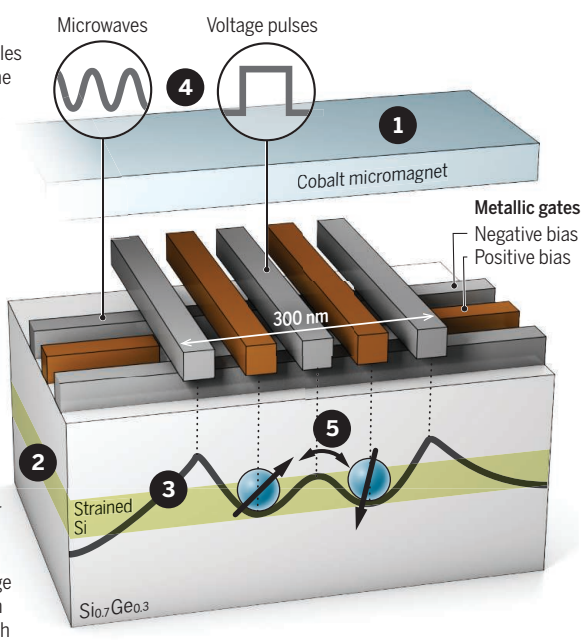
1 Stray field from micromagnet couples electron spins to the electric field.

2 Strain in Si layer vertically confines electrons.

3 Potential created by metal gates laterally confines electrons.

4 Spin alignment of the electrons is controlled with microwaves and voltage pulses.

5 Tunnel coupling through the barrier between the two electrons (blue) causes an exchange interaction between the two spins, which drives entanglement.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

REFERENCES

1. D. M. Zajac *et al.*, *Science* **359**, 439 (2018).
2. E. A. Chekhovich *et al.*, *Nat. Mater.* **12**, 494 (2013).
3. J. T. Muhonen *et al.*, *Nat. Nanotechnol.* **9**, 986 (2014).
4. D. M. Zajac *et al.*, *Phys. Rev. Appl.* **6**, 054013 (2016).
5. J. Yoneda *et al.*, arXiv:1708.01454 [cond-mat.mes-hall] (4 August 2017).
6. J. P. Dehollain *et al.*, *New J. Phys.* **18**, 103018 (2016).
7. M. Veldhorst *et al.*, *Nature* **526**, 410 (2015).
8. T. F. Watson *et al.*, arXiv:1708.04214 [cond-mat.mes-hall] (14 August 2017).
9. P. Cerfontaine *et al.*, *Phys. Rev. Lett.* **113**, 150501 (2014).
10. L. M. K. Vandersypen *et al.*, *npj Quant. Inf.* **3**, 34 (2017).
11. X. Mi *et al.*, arXiv:1710.03265 [cond-mat.mes-hall] (9 October 2017).
12. N. Samkharadze *et al.*, arXiv:1711.02040 [cond-mat.mes-hall] (6 November 2017).

10.1126/science.aar6209

POLICY FORUM

ENERGY AND THE ENVIRONMENT

Engaging over data on fracking and water quality

Data alone aren't the solution, but they bring people together

By **S. L. Brantley,[1,2] R. D. Vidic,[3] K. Brasier,[4] D. Yoxtheimer,[5] J. Pollak,[6] C. Wilderman,[7] T. Wen[1]**

Extraction of unconventional oil and gas using high-volume hydraulic fracturing (HVHF)—"fracking"—is a "wicked" problem: Science and policy-making are complex and opaque; problems are unstructured, cross areas of policy jurisdiction, require coordinated action among various stakeholders who disagree about values, and tend to result in limited solutions with complex consequences (*1*). Public participation in decision-making about hydrocarbon extraction is limited by the largely private nature of transactions among mineral rights owners and industry and the narrow opportunity for public input into procedures. Likewise, obstacles to accessing water-quality data and the dearth and diversity of such data limit shared understanding. We found, however, that, although data alone do not resolve wicked problems, shared interest in gathering, discussing, and improving water-quality data can lead to productive discussions among scientists, citizens with local knowledge, regulators, and industry practitioners. Such opportunities to "pull back the curtain" on science, funded and facilitated by honest brokers, could build trust and develop procedural fairness as foundations for social license.

The rate of HVHF in Pennsylvania (PA) increased exponentially from 2004 to 2011 because of rapid technological advances that accelerated development of the Marcellus formation, the largest such shale-gas play in the United States. Almost 11,000 shale-gas wells now dot the valleys and ridges across half of PA. Spills and leaks of fracking fluids and wastes occasionally occur, and although most are small, they add to the risk of cumulative impact. Events such as well blowouts and burning tap water amplified the public's perception of risks to water quality. Although many community-based watershed organizations monitor streams and are concerned about fracking, the state has struggled to build capacity to document pre-drilling water quality and postdrilling impacts.

Although some communication among scientists and nonscientists has focused on the impact of shale-gas development, this has generally targeted seismic risk rather than water quality. In 2011, we had the idea that assessing water-quality data in PA might help address public concerns if data from academics, consultants, industry practitioners, government, and nonscientists from watershed groups were compiled in one public, online database. Although the resulting Shale Network database itself has

Methane in groundwater at concentrations high enough to be flammable can occur naturally or be related to natural gas wells. Flaming tap water was found in Granville Summit, Pennsylvania, March 2012.

PHOTO: MELANIE STETSON FREEMAN/THE CHRISTIAN SCIENCE MONITOR VIA GETTY IMAGES

[1]*Earth and Environmental Systems Institute, Pennsylvania State University, University Park, PA, USA.* [2]*Department of Geosciences, Pennsylvania State University, University Park, PA, USA.* [3]*Department of Civil and Environmental Engineering, University of Pittsburgh, Pittsburgh, PA, USA.* [4]*Department of Agricultural Economics, Sociology, and Education, Pennsylvania State University, University Park, PA, USA.* [5]*Marcellus Center for Outreach and Research, Pennsylvania State University, University Park, PA 16802, USA.* [6]*Consortium of Universities for the Advancement of Hydrologic Sciences, Inc., Cambridge, MA 02140, USA.* [7]*Department of Environmental Studies, Dickinson College, Carlisle, PA, USA. Email: brantley@eesi.psu.edu*

proven useful, even more important has been the process of building it. We design workshops to engage data providers, recognizing that no one person understands all aspects of water data, from landscape to laboratory to computer to sharing to interpretation. These workshops forge a social network for volunteers, industry practitioners, consultants, regulators, and academic scientists voicing diverse perspectives and concerns about water quality, and the focus on data and observation keeps conversations productive. To our knowledge, no such network exists in other U.S. shale plays. The lessons we describe below from the effort cross the biophysical and social sciences, creating community among stakeholders.

Today, our database (*2*) spans ~28,000 sites mostly within PA and contains more than a million data values derived from multiple universities, government entities, volunteer groups, energy companies, and consultants (*3, 4*). All location-specific water-quality data with sufficient quality control can be published in the database, which is run collaboratively with CUAHSI (Consortium of Universities for the Advancement of Hydrologic Science, Inc.) as part of its National Science Foundation–funded work to provide a search engine that finds water data in online databases such as ours, as well as data from the U.S. Geological Survey (USGS), the U.S. Environmental Protection Agency (EPA), and universities and other entities.

Much of the surface-water data in the Shale Network database derive from sensors run by watershed groups, as well as from state agencies that had not previously published data online. Some industry data for PA—so-called "pre-drill" data that are shared with state regulators but are collected by industry before drilling to protect against potential liability—are now being uploaded, from tens of thousands of groundwater samples with up to 50 analytes each.

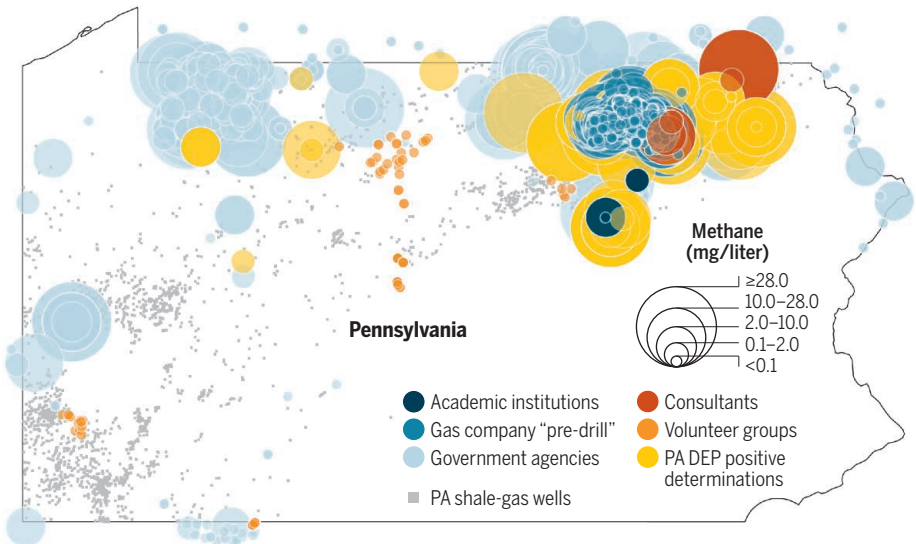
LESSONS ON WATER AND DATA

Given that water quality shows extreme spatial and temporal variability caused by both natural and anthropogenic factors, contamination from shale-gas activities can be difficult or impossible to document. Hundreds of HVHF chemicals are used that require diverse sampling and analytical strategies. In addition to HVHF chemicals, the wastewaters can carry naturally occurring radioactive materials, organics, metals, salts, and sediments (*5*). Toxic compounds that can transform in the environment are present at low, difficult-to-detect concentrations [e.g., (*6*)].

The constituents that are most likely to pollute waterways—sediments, salts, and methane—are already ubiquitous in the environment, derived from soils (sediments), oceanic aerosols, natural brines, road deicing (salts), and bacteria or rock formations (methane). In PA, methane also emits from many of the hundreds of thousands of mapped and unmapped legacy coal mines and gas or oil wells. Documenting new contamination in precontaminated waters is difficult (see the figure).

Shale-gas wells and water methane concentrations in Pennsylvania

Data showing locations of wells drilled since 2004 are from the Pennsylvania Department of Environmental Protection (PA DEP). Methane data are from the Shale Network database (*2*), contributed by sources as indicated in the legend. All methane data describe groundwater, except data from volunteer groups that measured surface water. PA DEP positive determinations indicate that the PA DEP investigated and concluded that a shale-gas company was presumed responsible for contamination. Concentrations above 10 mg/liter in drinking water wells are considered by the USGS as requiring action.



For many of the tens of thousands of stream kilometers and aquifer hectares near well pads in PA, there are insufficient water-quality data during relevant time periods and in the public domain to assess impacts. This is partly because the sources of potential contamination are widely distributed amid a complex stream network, aquifers are extremely heterogeneous, and contamination can be transient (*7*). Many of the measurements came from sensors located in limited locations, are focused on irrelevant analytes, or were only completed once, before or after drilling. Mostly, water quality is monitored in the United States in time-limited projects driven by concerns about specific phenomena (e.g., coal mining, acid rain).

Overall, we observed that stakeholders generally hesitate to share data (*3*). Volunteer groups lack expertise or time. Homeowners fear lower property resale values. Academics fear competition around publications and funding. Government and industry practitioners lack resources or express concerns about liability, controversy, and nondisclosure agreements. The result is that when papers are published without explicit data values and locations, findings are highly controversial (*8*).

Some Shale Network participants began to teach better data management for watershed groups (*4*), whereas others worked to broker the agreements that led to publication of "pre-drill" industry data. But data providers generally found it arduous to compile metadata to describe their measurements. Most preferred using their own databases. Water metadata are particularly complex in comparison to other monitored features of fracking such as seismicity because of the multitude of contaminants, names, and reporting conventions for contaminants, sampling strategies, and analytical methods. The common constituent nitrate is published in 13 different reporting conventions in PA data alone. Such issues also hamper new, innovative industry-government partnerships to publicize fracking chemicals (*9*).

Given these factors, it is not surprising that we know of <10 incidents where data in the database marginally document contamination effects through 2017. Neither has the database ever documented a previously unreported event. Instead, incidents are reported by the public, media, regulators, or industry. After discussing such observations at workshops, PA researchers began to focus on improving baseline stream chemistry estimates (*10*) or documenting cumulative impacts using sediments, isotopes, or

GRAPHIC: J. YOU/*SCIENCE*

ecological indicators (*11*). Workshop discussions also introduced new techniques to use gaining streams to monitor for methane in groundwater over wider areas (*12*).

These issues, as well as the difficulties and expense of monitoring, began to drive some conversations about coordinating plans for where and what to measure. But monitors tended to maintain local, expedient sampling designs. The Shale Network effort and the new search engine for water data nonetheless provide the first step in coordination by making data easier to find. Other monitoring groups are now looking at using the CUAHSI online data system. And the ongoing publication of large volumes of pre-drill data is driving conversations about better methods of data storage while highlighting anomalies in groundwater chemistry that might bear further investigation (*13*).

FACILITATING DATA-DRIVEN DIALOGUE

Several attributes of the workshops—some noted previously by social scientists studying other initiatives and issues—led to this success. First, workshops were facilitated by funding from an agency and universities identified as honest brokers (*14*), allowing the framework to be open and not driven by an agenda. No one pushed for decisions or consensus. Scientific analyses were presented that spoke to many policy options.

Second, all data with quality control were welcomed. We stressed participation by academics, government scientists, industry practitioners, consultants, and watershed groups. Everyone was encouraged to explore data in hands-on modules and to make oral or poster presentations. The focus was on observations emphasizing distinctive expert and local knowledge (*15*).

Third, we encouraged participants to articulate new foci for future workshops. Although motivations differed among stakeholders—scientists attended workshops to present or find data whereas nonscientists attended to understand issues in their backyards—conversations facilitated cooperative agenda setting (*15*) that built trust. In 2017, for example, a nonscientist expressed concerns about disposal practices for the ~1200 tons of rock fragments accumulated per gas well, and researchers now investigate where this rock is buried.

As understanding grew that the database did not reveal much evidence of contamination, we considered if the reason was lack of data or lack of incidents (*7*). Media and government reports for PA were scrutinized for incidents where wastes contaminated water resources through spills or leaks at the surface: In the 10-year period when ~10,000 wells were drilled after 2004, <100 spills or leaks greater than 400 gallons (*7*) transiently contaminated <200 km of the ~70,000 km of PA streams overlying the Marcellus play. Using the state online reports, the average distance between spills and the nearest stream was <300 m, and many of the streams had high ecological or drinking water value (*16*). Incidents were widely but unevenly distributed throughout the gas play and appeared more frequent early in the boom (*7, 16*). Fracking chemicals were never reported to migrate from deep shale into drinking water aquifers.

Given that these incidents largely could not be corroborated by data in our database, we began to work more closely with the state to learn about regulatory actions and difficulties. As reported by the state, methane migration was the most commonly reported problem: ~39 of the ~9600 shale wells drilled into the Marcellus in PA between 2004 and 2015 allowed methane migration, affecting 108 drinking water supplies. But some suspect that state reporting is limited by financial and human resource constraints. So, we began developing data mining tools to investigate both natural and non-natural sources of methane (*13*).

> ***"...recognition that groundwater is a public resource that requires public data transcends issues around shale gas..."***

When water-quality information for incidents was released publicly, we discussed the data (*6*). Discussions taught everyone about difficulties in determining causation, highlighting the need for multiple lines of evidence and state-of-the-art analyses. The public, regulators, academic scientists, consultants, and industry workers were together exposed to the knowns, unknowns, and gray areas.

Workshop participants remained concerned about the lack of disclosure when incidents were litigated and nondisclosure agreements signed. No regulations require data disclosure that could allow scientists in industry and academia to learn to improve best practices. In addition, no database of spill timing, volume, or cause is maintained in PA (*16*), and media and government reports of these incidents can be discrepant. These factors exacerbate our conundrum: How can we maintain public trust in water quality knowing that all data are not released and we cannot monitor everywhere a spill or leak might occur?

BARRIERS TO ENGAGEMENT

Data sharing can promote understanding and trust among stakeholders—but technical and nontechnical barriers around data must be surmounted. Water-quality investigators must follow the path that academic and USGS seismologists have followed by agreeing on standards for measuring and reporting.

Changing the norms for reporting is especially important for groundwater because, in many areas worldwide, including PA, groundwater is treated as private property yet moves across ownership boundaries. The recognition that groundwater is a public resource that requires public data transcends issues around shale gas and has already galvanized legislation in states experiencing drought, such as California.

Why doesn't data-driven engagement among scientists and nonscientists happen more often? After all, the American Petroleum Institute encourages community engagement early in exploration, and some gas companies participate in community programs. Rapid rates of technological change may outpace the rate of public engagement and policy implementation. Scientists often train to understand only select types of data and thus develop only fragmentary understanding, hampering communication with nonscientists. Limited funding from honest brokers hampers creation of unbiased forums for stakeholders to assess risks by publicly engaging around shared data. Forging opportunities to discuss science by focusing on data can nonetheless build trust and be part of the solution. ■

REFERENCES AND NOTES

1. E. P. Weber, A. M. Khademian, *Public Adm. Rev.* **68**, 334 (2008).
2. ShaleNetwork Database, DOI: 10.4211/his-data-shalenetwork.
3. K. J. Brasier *et al.*, *J. Environ. Plann. Manage.* **60**, 2103 (2017).
4. C. C. Wilderman, J. Monismith, *Citizen Science: Theory and Practice* **1**, 7 (2016).
5. R. D. Vidic, S. L. Brantley, J. M. Vandenbossche, D. Yoxtheimer, J. D. Abad, *Science* **340**, 826 (2013).
6. G. T. Llewellyn *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 6325 (2015).
7. S. L. Brantley *et al.*, *Int. J. Coal Geol.* **126**, 140 (2014).
8. S. G. Osborn, A. Vengosh, N. R. Warner, R. B. Jackson, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 8172 (2011).
9. K. Konschnik, *Nat. Resour. J.* **54**, 319 (2014).
10. X. Niu *et al.*, *Environ. Geochem. Health* 10.1007/s10653-017-0031-6 (2017).
11. W. D. Burgos *et al.*, *Environ. Sci. Technol.* **51**, 8851 (2017).
12. V. M. Heilweil *et al.*, *Environ. Sci. Technol.* **49**, 4057 (2015).
13. Z. Li *et al.*, Corrigendum to "Searching for anomalous methane in shallow groundwater near shale gas wells" [*J. Contam. Hydrol.* **195**, 23 (2016)]. *J. Contam. Hydrol.* **207**, 50 (2017).
14. R. A. Pielke Jr., *The Honest Broker: Making Sense of Science in Policy and Politics* (Cambridge Univ. Press, New York, 2007).
15. F. Fischer, *Citizens, Experts, and the Environment* (Duke Univ. Press, 2000).
16. K. O. Maloney *et al.*, *Sci. Total Environ.* **581-582**, 369 (2017).

ACKNOWLEDGMENTS

Funding derived from National Science Foundation grants OCE-11-40159, EAR-13-31726, and IIS-16-39150 to S.L.B.

10.1126/science.aan6520

Humans aren't alone in acknowledging death. Mourning has been noted in elephants, for example.

PSYCHOLOGY

Investigating the afterlife

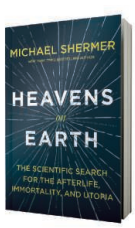
Faulty religious reasoning and sloppy secular arguments earn a skeptic's side-eye

By **Paula Quinon**

Awareness of one's mortality is universal for human beings. Its conceptualization often starts at a very young age, as preschoolers observe and note the differences between alive and dead organisms. Between the ages of 7 and 10, children come to understand that death is permanent and irreversible, which often leads to anxiety or a fear of death. Despite a growing ability to rationalize, this fear persists in adults.

Various cultural inventions aim to remedy this disturbing emotional reaction. Each transcendental explanation of what happens when we die is inevitably followed by an additional mythology that provides a cognitive reinforcement, resulting in a strong, coherent set of beliefs.

In his new book, *Heavens on Earth*, Michael Shermer aims to deconstruct systems of irrational beliefs. In particular, he provides analysis of three concepts: The first is a belief in immortality or in an afterlife. Sometimes this belief relates not to an individual but to the preservation of a species. The second is a belief in a utopia where a better version of one's self lives peacefully among peers. The third is a belief that one belongs to a group that has special insight into the nature of life after death. As Shermer shows, prognosticators of both religious and secular utopias can fall victim to this way of thinking.



Heavens on Earth
Michael Shermer
Henry Holt, 2018.
320 pp.

It is not difficult to imagine the evolutionary reasons behind these beliefs: A fear of death is a rationale for a belief in immortality; a desire for a well-functioning society provides foundations for utopias; a craving for meaning pushes people to perceive themselves as particularly important. Cognitive shortcuts that simplify complexity, such as confirmation bias (a tendency to search for information that is consistent with preexisting beliefs) or patternicity (a tendency to search for patterns in both meaningful information and meaningless noise), help to establish and reinforce this mode of thinking.

Throughout the book, Shermer investigates a multitude of examples, from religious heavens and spiritual traditions that search for universal consciousness and reincarnation to science-driven quests to create humans who will—in one form or another—live forever. He challenges each, in turn, using techniques from his skeptic's toolbox, reminding us, for example, that according to scientific methodology, a preponderance of anecdotes is not evidence for a preferred belief.

A skeptic, Shermer shows, must also be sensitive to the conceptual structure of the language used in argumentation. Techno-optimist theories based in science are a particularly rich area for this type of consideration. Shermer analyzes both transhumanists, who focus on preserving the physical body, and singularitarians, who, inspired by the rapid proliferation of information and computing technologies, believe that the self is simply a pattern of information that we will eventually be able to move from the biological brain to an artificial one.

Conceptual inconsistencies within these frameworks are frequent. For instance, if we want to combat aging, what do we define "aging" to mean? More fundamentally, how do we define self-identity? Shall I insist that it is inseparable from my body? If I upload my mind to a computer, how shall I conceive of my original self versus the copy?

Shermer also touches on the ways that strong beliefs can be used in various, not always commendable, manners. True, strong ideologies can motivate visionaries like Ray Kurzweil, Peter Thiel, or Elon Musk. But beliefs can also be used to encourage people to commit atrocities, as the Nazis did with the idea of racial supremacy in pre–World War II Germany and ISIS does today, promising its martyrs a paradisiacal life after their suicidal sacrifice.

Even if we can rationally reject each irrational argument in favor of an afterlife, Shermer acknowledges that belief will persist and that this makes it real, in a way. "Heavens above may or may not be real, but heavens on earth are, at least in the minds of those who believe in them," he writes. "In that sense, the empyrean realm of gods and heavens that resides in the brains of believers is as real as anything in the terrestrial kingdom."

We all seek to live a meaningful and purpose-driven life. But how do we do so without invoking an afterlife? "Through recognition of our uniqueness, through our gratitude for having the chance to live, through the love of others and others' love for us, and through engagement with the world with courage and integrity," writes Shermer. We can find "heavens on Earth," he argues, right here in the wonders of our own universe. ■

The reviewer is in the Department of Philosophy, Lund University, Lund, Sweden. Email: paula.quinon@fil.lu.se

10.1126/science.aar5147

PHOTO: MOF/ISTOCKPHOTO

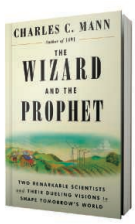
SCIENCE LIVES

How shall we save the planet?

A techno-optimist is pitted against the pied piper of "apocalyptic environmentalism"

By **Tyler Priest**

Charles Mann's fascinating new book, *The Wizard and the Prophet*, is about two pivotal but largely forgotten men of science, Norman Borlaug (1914–2009) and William Vogt (1902–1968). His inspiration, however, originated with two members of the opposite sex: his infant daughter and a former neighbor, the famous biologist and evolutionary theorist Lynn Margulis. "When my daughter is my age, almost 10 billion people will be walking the earth," he calculated. "How is *that* going to work?" The mischievous Margulis liked to compound Mann's worry by telling him that *Homo sapiens*, like other "briefly successful species," was biologically destined to multiply beyond its material limits before collapsing into extinction.



The Wizard and the Prophet
Charles C. Mann
Alfred A Knopf, 2018. 629 pp.

An inquisitive and gifted science writer, Mann is the author of previous bestsellers, *1491* and *1493*, which reappraised the ecological dimensions of human development in the Americas before Columbus and around the world after him. In *The Wizard and the Prophet*, Mann turns his gaze on the recent past and near future to see whether the human species can continue its improbable run of success. His chief subjects were visionaries "shouting from the edge of the petri dish" who had fundamentally different ideas about how to avert a planetary crisis.

The prophet Vogt, son of German immigrants living on Long Island and a self-taught ornithologist and zoologist, crusaded for population control and aggressive restraints on human consumption. He rose through the bird world to become editor of the Audubon Society's *Bird-Lore* in 1934 and later served as conservation chief of the Pan American Union and national director of the Planned Parenthood Federation of America. A confidant of Aldo Leopold and Julian Huxley, whose eugenicist tendencies he shared, Vogt vaulted to international fame with his 1948 bestseller, *Road to Survival*, in which he raged against "excessive breeding and abuse of the land." Humans, he warned, would soon "feel the chill of scarcity's damp breath." Mann casts Vogt as the central figure of midcentury "Malthusianism" and pied piper of "apocalyptic environmentalism," whose ideas influenced Rachel Carson and Paul Ehrlich.

The wizard Borlaug, grandson of Norwegian immigrants from the tiny Iowa farming town of Saude, was a nose-to-the-soil plant pathologist who believed that the application of science and technology could liberate humans from the shackles of nature. After World War II, he worked with the Rockefeller Foundation's Mexican Agricultural Program to crossbreed wheat varieties that could resist stem rust and boost yields. Years of painstaking experiments eventually produced enough wheat from new superseeds to fend off starvation in Mexico, India, and other impoverished parts of the world. Borlaug's starring role in the "Green Revolution," which was widely credited with saving tens of millions of lives, earned him the Nobel Peace Prize in 1970.

Mann tells the stories of Vogt and Borlaug with verve and suspense. The two men met each other once, briefly, in Mexico in 1946, and neither was impressed with the other. Upon seeing what Borlaug was up to, Vogt urged the Rockefeller Foundation to halt it.

Particularly revealing is the poignant tale of Vogt's 4 years (1938–1942) on the desolate and noxious guano islands off Peru studying the declining populations of the guanay cormorant, an experience that reinforced his conviction that all species were bound by hard ecological limits. Equally engrossing are Borlaug's desperate maneuvers in 1965 to ship more than 500 tons of high-yielding seeds from Mexico through Los Angeles during the upheaval of the Watts riots and then on to India and Pakistan, which were at war with each other over Kashmir.

The achievement of *The Wizard and the Prophet* goes beyond detailing the accomplishments of these two individuals. With his trademark interdisciplinary wisdom, Mann travels the globe to visit other wizards and prophets working on the world's most challenging environmental problems: organic farming versus genetic manipulation in a section on "Earth: Food"; sewage recycling versus aqueducts and desalination in "Water: Freshwater"; centralized versus decentralized forms of energy delivery in "Fire: Energy"; and the race between decarbonization and geoengineering to fight global warming in "Air: Climate Change."

The visions of Vogt and Borlaug do not have to be mutually exclusive, but most solutions to our toughest dilemmas lean toward one or the other. So where does Mann stand? Readers will have to interpret his conclusion for themselves.

After describing the progress modern humans have made in formally abolishing slavery, alleviating the subjugation of women, and reducing violent death and deprivation, Mann shudders to think that we "have the imagination to see our potential end, but [do] not have the cultural resources to avoid it." The power to do both is what distinguishes us from the "protozoa in the petri dish." ■

10.1126/science.aar2447

The reviewer is in the Department of History, University of Iowa, Iowa City, IA 52242, USA. Email: tyler-priest@uiowa.edu

PHOTO: PA IMAGES/ALAMY STOCK PHOTO

Lab-grown burgers, like this one created by Mark Post at Maastricht University, may one day be the norm.

PODCAST

Clean Meat
How Growing Meat Without Animals Will Revolutionize Dinner and the World
Paul Shapiro
Gallery Books, 2018. 256 pp.

As the demand for meat increases and concerns about industrial animal agriculture mount, the need to create a safer, more efficient system of meat production has become urgent. This week on the *Science* podcast, Paul Shapiro describes how tiny tissue samples may one day be used to generate healthy, humanely produced meat. sciencemag.org/podcasts

10.1126/science.aas8716

Conscious robots may merit legal protections.

PHOTO: AP PHOTO/IRAM ASIM

LETTERS

Edited by **Jennifer Sills**

Conscious machines: Defining questions

In their Review "What is consciousness, and could machines have it?" (27 October 2017, p. 486), S. Dehaene *et al.* argue that the science of consciousness indicates that we are not on the verge of creating conscious machines. However, Dehaene *et al.* ask and answer the wrong questions.

To determine whether machines are conscious, we must ask whether they have subjective experiences: Do machines consciously perceive and sense colors, sounds, and smells? Do they feel emotions? Unfortunately, Dehaene *et al.* relegate this issue to the final paragraph of their Review, dismissing it as a philosophical question "beyond the scope of the present paper." Instead, they ask whether machines "mimic" consciousness by exhibiting the global availability of information (the ability to select, access, and report information) and metacognition (the capacity for self-monitoring and confidence estimation). Questions concerning to what extent machines have these capacities are interesting, but neither capacity is necessary or sufficient for subjective experience (*1*, *2*). Furthermore, Dehaene *et al.*'s emphasis on metacognition and global broadcasting presumes that the prefrontal cortex is the home of consciousness, which remains a matter of debate (*3*, *4*).

Finally, when arguing that machines are not yet conscious, Dehaene *et al.* highlight the "feedforward"—i.e., sequential—nature of information processing typical of these computing systems. Although current machines may not exhibit propagation of information into the broadcasting network hub or formal Bayesian metacognition, many artificial intelligence systems involve global projection of winning representations and have metacognitive capacity, such as confidence estimates. If these types of processes were strongly indicative of consciousness, we would have to admit that some machines are already conscious.

We agree with Dehaene *et al.* that to address the question of machine consciousness, we must start with a theory of human consciousness. Given current disagreement on that topic, we are all left to speculate whether machines will ever be conscious. A more pertinent question for the field might be: "What would constitute successful demonstration of artificial consciousness?"

Olivia Carter,[1] Jakob Hohwy,[2] Jeroen van Boxtel,[2] Victor Lamme,[3] Ned Block,[4] Christof Koch,[5] Naotsugu Tsuchiya[2]*
[1]University of Melbourne, Melbourne, VIC 3010, Australia. [2]Monash University, Mebourne, VIC 3800 Australia. [3]University of Amsterdam, 1018 XA Amsterdam, Netherlands. [4]New York University, New York, NY 10003, USA. [5]Allen Institute for Brain Science, Seattle, WA 98103, USA.
*Corresponding author.
Email: naotsugu.tsuchiya@monash.edu

REFERENCES

1. C. Koch *et al.*, *Nat. Rev. Neurosci.* **17**, 307 (2016).
2. N. Tsuchiya *et al.*, *Trends Cogn. Sci.* **19**, 757 (2015).
3. M. Boly *et al.*, *J. Neurosci.* **37**, 9603 (2017).
4. B. Odegaard, R. T. Knight, H. Lau, *J. Neurosci.* **37**, 9593 (2017).

10.1126/science.aar4163

Conscious machines: Robot rights

In their Review "What is consciousness, and could machines have it?" (27 October 2017, p. 486), S. Dehaene *et al.* suggest that machine consciousness, which would model human cognitive functions within a physical architecture other than the human brain, has not yet been achieved. Creating consciousness is still a goal for the future, but now is the time to consider the implications of conscious machines.

Developing machine capacities such as artificial intelligence (AI) and robots for human-robot interaction is of extreme technological value, but it raises moral and ethical questions. For example, one of the major sectors in robotics is the development of sexual robots (*1*). Although some researchers consider this phenomenon a gateway to the future acceptance of human-robot interactions, others see a danger for human society as robots modify the social representation of human behaviors (*2*, *3*).

The robotics field must wrestle with these questions: Is it ethical to create and continue to use robots with consciousness in the way we use robots that were originally designed for our needs? Do robots deserve to be protected? This debate could challenge the limits of human morality and polarize society's views on whether conscious robots are objects or living entities (*4*–*6*).

As we approach an era when conscious robots become part of daily life, it is important to start thinking about the current status of robots today. The purpose of granting legal status to robots is not only to prevent inappropriate human-robot interactions but also to recognize and formalize the role of robots in society, thus normalizing their existence (*7*).

Nicolas Spatola* and Karolina Urbanska
Université Clermont Auvergne, CNRS, LAPSCO, F-63000 Clermont-Ferrand, France.
*Corresponding author. Email: nicolas.spatola@uca.fr

REFERENCES

1. M. Scheutz, T. Arnold, "Intimacy, bonding, and sex robots: Examining empirical results and exploring ethical ramifications" (2017); https://hrilab.tufts.edu/publications/scheutz2017intimacy.pdf.
2. J. Robertson, *Body Soc.* **16**, 1 (2010).
3. M. Coeckelbergh, *Int. J. Soc Robot.* **1**, 217 (2009).
4. P. Lin *et al.*, Eds., *Robot Ethics* (MIT Press, 2012).
5. A. L. Peláez, D. Kyriakou, *Technol. Forecast. Soc. Change* **75**, 1176 (2008).
6. K. Richardson, *Comp. Soc.* **45**, 290 (2016).
7. I. Yeoman, M. Mars, *Futures* **44**, 365 (2012).

10.1126/science.aar5059

Response

Any discussion of machine consciousness should start with empirical evidence, and our Review primarily consists of an empirical look at how nonconscious and conscious processing differ in humans [see also (*1*, *2*)]. In contrast to Carter *et al.*'s interpretation of our conclusions, we suggest that conscious subjective states are in fact on the verge of becoming implementable in machines and that two computational ingredients (global information sharing and self-monitoring), if jointly and correctly implemented, may provide machines with conscious subjectivity.

Carter *et al.* claim that "neither capacity is necessary or sufficient for subjective experience," but that is begging the question. It is possible that the subjective experiences of humans are simply information-bearing representations with the same specific properties we used for those of machines: being globally available and therefore reportable, and entering into dedicated self-monitoring processes capable of evaluation and criticism. Empirical evidence suggests that whenever human subjective experience is impaired, such as in psychosis (*3*) or blindsight (*4*), aspects of both of these processing functions are also disrupted. Thus, our working hypothesis is that subjective experience comes down to nothing but a combination of specific forms of processing (including reality monitoring as a crucial component).

It has been suggested that failing to recognize that machines may have subjective states reflects a lack of imagination (*5*). In fact, many if not all visual illusions, in which subjective perception diverges from objective reality, arise from efficient computing, such as applying a Bayesian prior to noisy inputs (*6*) or taking efficient shortcuts in otherwise intractable computations (*7*). Thus, such subjective percepts would arise in any efficient machine.

To move forward, Carter *et al.* point out that it would be advantageous if we could agree on the criteria that demonstrate consciousness. But it is precisely because we lack such a consensus that we think it is best to start with empirical evidence based on the known features of the human brain. Rapid progress in mapping the human brain mechanisms of consciousness has indeed revealed an important contribution of the prefrontal cortex [e.g., (*8*–*10*)].

When machines share enough features with conscious human brain processing, we should be prepared to accept the possibility that they are conscious. Of course, even if a machine shared those features and reported having subjective experiences, Carter *et al.* could still deny that it experienced anything at all. But such a solipsist position also applies to humans: By such a standard, we likewise cannot prove that other human beings are conscious. This position, and the associated insistence on "qualia" and the "hard problem" of consciousness, are unproductive (*11*). In the future, denying machines any form of subjectivity, when it is caused by computations similar to those that constitute core ingredients of consciousness in the human brain, may become as contentious as denying it to other human beings or to nonhuman animals with brain architectures similar to ours.

We therefore agree with Spatola and Urbanska that the predictable emergence of conscious machines calls for an immediate consideration of its societal consequences. The potential benefits should not be neglected: A powerful sentient artificial intelligence (AI) may collaborate with humans in addressing major issues such as energy management, ecology, or care in an aging society. The risks, however, are equally real and include job loss, concentration of power in a few hands, a military arms race, and social upheaval as humans and AI increasingly compete for the same societal roles. Mitigating these disorders will require a major international effort, and we can only heed here the conclusion of a recent academic statement on the power and limits of AI (*12*): "Just like crash tests for transportation, the passing

of ethical and safety tests, evaluating, for instance, social impact or racial prejudice, could become a prerequisite to the release of [artificial intelligence] software."

Stanislas Dehaene,[1,2]* Hakwan Lau,[3,4] Sid Kouider[5]

[1]Chair of Experimental Cognitive Psychology, Collège de France, 75005 Paris, France. [2]Cognitive Neuroimaging Unit, Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), INSERM, Université Paris-Sud, Université Paris-Saclay, NeuroSpin Center, 91191 Gif/Yvette, France. [3]Department of Psychology and Brain Research Institute, University of California, Los Angeles, Los Angeles, CA 90095, USA. [4]Department of Psychology, University of Hong Kong, Hong Kong. [5]Brain and Consciousness Group (École Normale Supérieure, École des Hautes Études en Sciences Sociales, CNRS), Département d'Études Cognitives, École Normale Supérieure–Paris Sciences et Lettres Research University, Paris, France.
*Corresponding author.
Email: stanislas.dehaene@cea.fr

REFERENCES

1. S. Kouider, S. Dehaene, *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **362**, 857 (2007).
2. S. Dehaene, *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts* (Penguin Books, Reprint edition, 2014).
3. P. C. Fletcher, C. D. Frith, *Nat. Rev. Neurosci.* **10**, 48 (2009).
4. Y. Ko, H. Lau, *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **367**, 1401 (2012).
5. D. Dennett, *Cognition* **79**, 221 (2001).
6. Y. Weiss *et al.*, *Nat. Neurosci.* **5**, 598 (2002).
7. T. D. Ullman, E. Spelke, P. Battaglia, J. B. Tenenbaum, *Trends Cogn. Sci.* 10.1016/j.tics.2017.05.012 (2017).
8. M. Wang, D. Arteaga, B. J. He, *Proc. Natl. Acad. Sci. U.S.A.* **110**, E3350 (2013).
9. T. I. Panagiotaropoulos *et al.*, *Neuron* **74**, 924 (2012).
10. B. Odegaard, R. T. Knight, H. Lau, *J. Neurosci. Off. J. Soc. Neurosci.* **37**, 9593 (2017).
11. D. C. Dennett, in *Consciousness in Modern Science*, A. Marcel, E. Bisiach, Eds. (Oxford University Press, 1988), pp. 42–77.
12. Pontifical Academy of Sciences, "Final statement of the workshop: Power and limits of artificial intelligence" (2016); www.pas.va/content/accademia/en/events/2016/intelligence/statement.html.

10.1126/science.aar8639

TECHNICAL COMMENT ABSTRACTS

Comment on "Precipitation drives global variation in natural selection"

Isla H. Myers-Smith and Judith H. Myers

Siepielski *et al*. (Reports, 3 March 2017, p. 959) claim that "precipitation drives global variation in natural selection." This conclusion is based on a meta-analysis of the relationship between climate variables and natural selection measured in wild populations of invertebrates, plants, and vertebrates. Three aspects of this analysis cause concern: (i) lack of within-year climate variables, (ii) low and variable estimates of covariance relationships across taxa, and (iii) a lack of mechanistic explanations for the patterns observed; association is not causation.
Full text: dx.doi.org/10.1126/science.aan5028

Response to Comment on "Precipitation drives global variation in natural selection"

Adam M. Siepielski, Michael B. Morrissey, Mathieu Buoro, Stephanie M. Carlson, Christina M. Caruso, Sonya M. Clegg, Tim Coulson, Joseph DiBattista, Kiyoko M. Gotanda, Clinton D. Francis, Joe Hereford, Joel G. Kingsolver, Kate E. Augustine, Loeske E. B. Kruuk, Ryan A. Martin, Ben C. Sheldon, Nina Sletvold, Erik I. Svensson, Michael J. Wade, Andrew D. C. MacColl

The Comment by Myers-Smith and Myers focuses on three main points: (i) the lack of a mechanistic explanation for climate-selection relationships; (ii) the appropriateness of the climate data used in our analysis; and (iii) our focus on estimating climate-selection relationships across (rather than within) taxonomic groups. We address these critiques in our response.
Full text: dx.doi.org/10.1126/science.aan5760

U.S. policy on Cuba obstructs crucial medical breakthroughs

AAAS Cuban Biomedical Fellows delayed in obtaining visas for collaborations on Alzheimer's, cancer

By **Michaela Jarvis**

The new U.S. policy on Cuba is holding back science collaborations that can accelerate by years the development of early detection and treatment technologies for such diseases as Alzheimer's and cancer.

In July 2017, before the new policy was imposed, Cuban neuroscientist Marquiza Sablón traveled to Washington University in St. Louis (WUSTL) as a participant in the AAAS Cuban Biomedical Fellows program. At her home institution, the Cuban Neurosciences Center, Sablón had developed compounds with the potential to detect Alzheimer's disease when tagged with radioactive isotopes, a promising new technology that could potentially be used for diagnosis and early detection based on nuclear imaging techniques. By the time of her departure at the end of December, Sablón said her "research goal was reached with the use of a radionuclide produced at WUSTL," and she and her U.S. colleagues hoped to publish an article on the research to share their findings with other pioneers in the field.

"Through this six-month exchange," said Buck Rogers, director of the lab that hosted Sablón at WUSTL, "Marquiza was able to make progress and answer questions about her project that may have taken years in Cuba."

"The whole lab learned from her experience and unique perspective while working on the experiments associated with her project," Rogers added. "From my experience, it has been a highly productive exchange that benefits both the U.S. and Cuban labs. It's always beneficial to have new insights and broader perspectives contributing to tough questions."

Other Cuban researchers scheduled to work in U.S. labs through the AAAS Cuban Biomedical Fellows program are unfortunately blocked from obtaining visas in Cuba, after a change in U.S. policy occurred in late September with the hardening stance of President Trump toward the island nation and reports by U.S. embassy officials in Havana of having suffered mysterious health problems of unknown origin.

The AAAS program to bring Cuban medical researchers to work alongside their U.S. counterparts grew out of a long-standing history of scientific collaboration between the two nations and an agreement signed by AAAS and the Cuban Academy in 2014 to work together despite political obstacles on research important to the populations of both countries.

Darel Martínez is a researcher at Cuba's Center of Molecular Immunology, which has pioneered cancer antibodies and vaccines approved in several countries around the world. At the end of 2016, he was chosen to be a AAAS Cuban Biomedical Fellow in the lab of U.S. immunotherapy expert Carl June at the University of Pennsylvania to work on the design of a chimeric antigen receptor T cell to target a protein that plays a role in tumor development.

As a researcher, Martínez is a strong advocate of international collaboration, indicating that without it, scientific research is needlessly repeated and, in some cases, produces results that are no longer relevant. "The only way that science can progress constantly is through the accumulation of knowledge used as a basis to build the next level," Martínez said. "Without collaboration among scientific groups that show more progress in a particular field, the progress of the research slows, and probably when you reach your goals, they will be obsolete and useless."

Agreeing, Avery Posey, associate director of the June Lab, explained that Martínez has an expertise in the biology and antibody targeting of epidermal growth factor receptors, the main technology behind his institution's cancer vaccine. The Center for Cellular Immunotherapies at the University of Pennsylvania, meanwhile, has expertise in the development of T cell immunotherapies that are built from parts of therapeutic antibodies. "We are not sure what the outcomes will be of this collaboration," said Posey, "but one hope is that Cuba can figure out ways to drive down the manufacturing costs of CAR T cells." CAR T cell therapy can cost cancer

A Cuban flag flies outside the U.S. embassy in Havana.

PHOTO: ALEXANDRE MENEGHINI/REUTERS/NEWSCOM

patients upward of $350,000 for a one-time dose.

"International collaboration is essential," said Posey, "to ensuring that scientists develop and drive forward the best possible therapeutics and ideas, not just the ones that exist within our borders."

Unfortunately, Martínez has been thwarted from coming to collaborate with June's lab. After working through "a mountain of emails, immigration paperwork, and contracts," said Posey, Martínez received immigration documents and the lab paid fees for a visa interview in September. He was waiting for that interview to take place at the U.S. embassy in Havana when Hurricane Irma hit Cuba, and the consular section of the embassy was closed. When it reopened, the U.S. policy prohibiting issuing visas to Cuban citizens in Cuba had been instituted. On instructions from the U.S. State Department, Martínez must travel to a visa interview in a third country to enter the United States, and lab officials are hoping he will arrive in February.

"This process, which is not yet finished," said Martínez, "results in a delay of at least four months at the start of the collaboration project and implies an increase in the cost of my trip, money that could be better used on research."

"While not explicitly targeting science," said Marga Gual Soler, project director at the AAAS Center for Science Diplomacy, "this new U.S. policy on Cuba severely limits travel and the logistical aspects of scientific cooperation. These effects will likely disincentivize joint research on both sides."

Cuban cancer researchers Tays Hernández and Janoi Chang have also been blocked in their attempt to participate in the AAAS Cuban Biomedical Fellows program. The two were to have arrived in Boston by early this year to work with researchers at Boston Children's Hospital, using molecules made in Cuba that show promise in the development of new drugs and imaging tools for cancer therapy and autoimmune or inflammatory diseases. They have been unable, however, to even begin the visa application process.

"We couldn't travel to the United States, despite the financial support we had from AAAS," Hernández and Chang wrote in an email, referring to funding provided by the Lounsbery Foundation to the AAAS program. "For the moment, the project is blocked."

Sablón at Washington University, where she and U.S. colleagues worked on Alzheimer's detection techniques.

PHOTO: NILANTHA BANDARA, PHD

Results of the 2017 election of AAAS officers

Terms begin on 20 February 2018. The constitutional change increasing the number of Board-appointed members of the Committee on Nominations from two to four was approved.

GENERAL ELECTION

President-Elect: *Steven Chu, Stanford University*

Board of Directors: *Rosina M. Bierbaum, University of Michigan/University of Maryland; Laura H. Greene, National High Magnetic Field Laboratory/Florida State University*

Committee on Nominations: *Sallie (Penny) W. Chisholm, Massachusetts Institute of Technology; Mary Frank Fox, Georgia Institute of Technology; Jeffrey I. Gordon, Washington University School of Medicine in St. Louis; Amita Sehgal, University of Pennsylvania Perelman School of Medicine*

SECTION ELECTIONS

Agriculture, Food, and Renewable Resources

Chair Elect: *Alison L. Van Eenennaam, University of California, Davis*

Member-at-Large of the Section Committee: *Karen E. Koch, University of Florida*

Electorate Nominating Committee: *Vivek Kapur, Pennsylvania State University; Marcia M. Miller, Beckman Research Institute of City of Hope*

Council Delegate: *Jennifer L. Tank, University of Notre Dame*

Anthropology

Chair Elect: *Susan C. Antón, New York University*

Member-at-Large of the Section Committee: *Kaye E. Reed, Arizona State University*

Electorate Nominating Committee: *Daniel H. Sandweiss, University of Maine; Denise F. Su, Cleveland Museum of Natural History*

Astronomy

Chair Elect: *Dennis Zaritsky, University of Arizona*

Member-at-Large of the Section Committee: *Lynne A. Hillenbrand, California Institute of Technology*

Electorate Nominating Committee: *Tie (run-off election will be conducted)*

Atmospheric and Hydrospheric Sciences

Chair Elect: *Ana P. Barros, Duke University*

Member-at-Large of the Section Committee: *Stephanie Pfirman, Barnard College/ Columbia University*

Electorate Nominating Committee: *Meredith Hastings, Brown University; Cora E. Randall, University of Colorado*

Biological Sciences

Chair Elect: *Vicki L. Chandler, Minerva Schools at Keck Graduate Institute*

Member-at-Large of the Section Committee: *Carol L. Boggs, University of South Carolina*

Electorate Nominating Committee: *Bruce Bowerman, University of Oregon; Scott V. Edwards, Harvard University*

Council Delegates: *Elizabeth S. Haswell, Washington University in St. Louis; Catherine M. Pringle, University of Georgia; Alejandro Sánchez Alvarado, HHMI/ Stowers Institute for Medical Research*

Chemistry

Chair Elect: *Vicki H. Grassian, University of California, San Diego*

Member-at-Large of the Section Committee: *Jacquelyn Gervay-Hague, University of California, Davis*

Electorate Nominating Committee: *Jaqueline L. Kiplinger, Los Alamos National Laboratory; Elsa Reichmanis, Georgia Institute of Technology*

Council Delegate: *Teresa Fryberger, National Academies of Sciences, Engineering and Medicine*

Dentistry and Oral Health Sciences

Chair Elect: *Jacques Eduardo Nör, University of Michigan*

Member-at-Large of the Section Committee: *Anne George, University of Illinois at Chicago*

Electorate Nominating Committee: *Effie Ioannidou, UConn Health; Purnima Kumar, Ohio State University*

Education

Chair Elect: *Patricia Marsteller, Emory University*

Member-at-Large of the Section Committee: *Angela Calabrese Barton, Michigan State University*

Electorate Nominating Committee: *Stacey Kiser, Lane Community College; Kathrin F. Stanger-Hall, University of Georgia*

Engineering

Chair Elect: *Sarah A. Rajala, Iowa State University*

Member-at-Large of the Section Committee: *Ilona Kretzschmar, The City College of New York, CUNY*

Electorate Nominating Committee: *Christina Chan, Michigan State University; Eric K. Lin, National Institute of Standards and Technology*

Council Delegate: *Sankar Basu, National Science Foundation*

General Interest in Science and Engineering

Chair Elect: *Patricia L. Ward, Museum of Science and Industry, Chicago*

Member-at-Large of the Section Committee: *Sharon M. Friedman, Lehigh University*

Electorate Nominating Committee: *Susan Gaidos, Freelance; Mari N. Jensen, University of Arizona*

Geology and Geography

Chair Elect: *Paul L. Koch, University of California, Santa Cruz*

Member-at-Large of the Section Committee: *Emi Ito, University of Minnesota*

Electorate Nominating Committee: *Mary Anne Holmes, University of Nebraska-Lincoln (Retired); Sally E. Walker, University of Georgia*

History and Philosophy of Science

Chair Elect: *Michael R. Dietrich, University of Pittsburgh*

Member-at-Large of the Section Committee: *Judy Johns Schloegel, Independent Scholar*

Electorate Nominating Committee: *Sally Gregory Kohlstedt, University of Minnesota; Zuoyue Wang, California State Polytechnic University, Pomona*

Council Delegate: *James R. Fleming, Colby College*

Industrial Science and Technology

Chair Elect: *Martin Keller, National Renewable Energy Laboratory*

Member-at-Large of the Section Committee: *J. Stephen (Steve) Rottler, Sandia National Laboratory (Retired)*

Electorate Nominating Committee: *Philip M. Harrington, BASF Corporation; Anna W. Topol, IBM Research*

Council Delegate: *Cammy R. Abernathy, University of Florida*

Information, Computing, and Communication

Chair Elect: *Joseph Halpern, Cornell University*

Member-at-Large of the Section Committee: *Maja J. Matarić, University of Southern California*

Electorate Nominating Committee: *Lydia E. Kavraki, Rice University; Diane L. Souvaine, Tufts University*

Linguistics and Language Science

Chair Elect: *Tie (run-off election will be conducted)*

Member-at-Large of the Section Committee: *Keren Rice, University of Toronto (Canada)*

Electorate Nominating Committee: *William Badecker, National Science Foundation; Donca Steriade, Massachusetts Institute of Technology*

Mathematics

Chair Elect: *James M. Crowley, Society for Industrial and Applied Mathematics (SIAM)*

Member-at-Large of the Section Committee: *Irene Fonseca, Carnegie Mellon University*

Electorate Nominating Committee: *Suzanne Lenhart, University of Tennessee; Audrey Terras, University of California, San Diego*

Medical Sciences

Chair Elect: *Gilda A. Barabino, The City College of New York, CUNY*

Member-at-Large of the Section Committee: *Katherine Amberson Hajjar, Weill Cornell Medicine*

Electorate Nominating Committee: *Lisa M. Coussens, Oregon Health & Science University; Cathy L. Drennan, Howard Hughes Medical Institute/ Massachusetts Institute of Technology*

Council Delegate: *Stephen B. Baylin, Johns Hopkins University School of Medicine; Vivian G. Cheung, Howard Hughes Medical Institute/ University of Michigan*

Neuroscience

Chair Elect: *Marina R. Picciotto, Yale University*

Member-at-Large of the Section Committee: *Margaret M. McCarthy, University of Maryland School of Medicine*

Electorate Nominating Committee: *Anne C. Hart, Brown University; George (Ron) Mangun, University of California, Davis*

Pharmaceutical Sciences

Chair Elect: *Craig W. Lindsley, Vanderbilt University*

Member-at-Large of the Section Committee: *Amy M. Barrios, University of Utah*

Electorate Nominating Committee: *John A. Beutler, National Cancer Institute/National Institutes of Health; Jorge A. Iñiguez-Lluhí, University of Michigan*

Physics

Chair Elect: *Jonathan A. Bagger, TRIUMF/University of British Columbia (Canada)*

Member-at-Large of the Section Committee: *Tim M. P. Tait, University of California, Irvine*

Electorate Nominating Committee: *Nicholas P. Bigelow, University of Rochester; Myriam P. Sarachik, The City College of New York, CUNY*

Council Delegate: *Alan R. Bishop, Los Alamos National Laboratory*

Psychology

Chair Elect: *Susan A. Gelman, University of Michigan*

Member-at-Large of the Section Committee: *Kent Berridge, University of Michigan*

Electorate Nominating Committee: *Jennifer Crocker, Ohio State University; Mary Hegarty, University of California, Santa Barbara*

Council Delegate: *James L. McClelland, Stanford University*

Social, Economic, and Political Sciences

Chair Elect: *Barbara Schneider, Michigan State University*

Member-at-Large of the Section Committee: *Pamela E. Oliver, University of Wisconsin – Madison*

Electorate Nominating Committee: *Joane Nagel, University of Kansas; Elaine Wethington, Cornell University*

Council Delegate: *Karen S. Cook, Stanford University*

Societal Impacts of Science and Engineering

Chair Elect: *Kei Koizumi, American Association for the Advancement of Science*

Member-at-Large of the Section Committee: *Barbara Belmont, California State University, Dominguez Hills*

Electorate Nominating Committee: *Susan E. Cozzens, Georgia Institute of Technology; Jennifer Kuzma, North Carolina State University*

Statistics

Chair Elect: *M. Elizabeth Halloran, Fred Hutchinson Cancer Research Center/University of Washington*

Member-at-Large of the Section Committee: *Hongzhe Li, University of Pennsylvania*

Electorate Nominating Committee: *David Banks, Duke University; Jessica Utts, University of California, Irvine*

Call for nomination of 2018 Fellows

Fellows who are current members of AAAS are invited to nominate members for election as Fellows. A member whose efforts on behalf of the advancement of science or its applications are scientifically or socially distinguished, and who has been a continuous member for the 4-year period leading up to the year of nomination, may by virtue of such meritorious contribution be elected a Fellow by the AAAS Council.

A nomination must be sponsored by three previously elected AAAS Fellows (who are current in their membership), two of whom must have no affiliation with the nominee's institution. Nominations undergo review by the steering groups of the Association's sections (the chair, chair-elect, retiring chair, secretary, and four members-at-large of each section). Each steering group reviews only those nominations designated for its section. Names of Fellow nominees who are approved by the steering groups are presented to the Council in the fall for election.

Nominations with complete documentation must be received by 11 April 2018. Nominations received after that date or nominations that are incomplete as of the deadline will not move forward. Complete instructions and a copy of the nomination form are available at www.aaas.org/current-nomination-cycle. Questions may be directed to fellownomination@aaas.org.

RESEARCH

Ancient teeth suggest early human migration from Africa

Hershkovitz et al., *p. 456*

IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

Plastic waste is sickening corals.

CORAL REEFS

Corals wrapped in plastic

Coral reefs provide vital fisheries and coastal defense, and they urgently need protection from the damaging effects of plastic waste. Lamb *et al.* surveyed 159 coral reefs in the Asia-Pacific region. Billions of plastic items were entangled in the reefs. The more spikey the coral species, the more likely they were to snag plastic. Disease likelihood increased 20-fold once a coral was draped in plastic. Plastic debris stresses coral through light deprivation, toxin release, and anoxia, giving pathogens a foothold for invasion. —CA

Science, this issue p. 460

MEMBRANE TARGETING

A new way into the ER

Membrane-embedded proteins are highly diverse in topology, physical characteristics, and location. This diversity necessitates multiple pathways for their effective membrane insertion. Guna *et al.* found that a widely conserved protein complex is responsible for inserting a subset of membrane proteins into the endoplasmic reticulum (ER) membrane (see the Perspective by Fry and Clemons Jr.). This ER membrane protein complex (EMC) inserts transmembrane domains whose topology and hydrophobicity preclude effective recognition by other insertion factors. This finding helps explain why the loss of EMC causes ER stress and altered protein trafficking. —SMH

Science, this issue p. 470; see also p. 390

QUANTUM GASES

Tuning the atomic pairing

Cold atomic gases are extremely flexible systems; the ability to tune interactions between fermionic atoms can, for example, cause the gas to undergo a crossover from weakly interacting fermions to weakly interacting bosons via a strongly interacting unitary regime. Murthy *et al.* studied this crossover in a gas of fermions confined to two dimensions. The formation of atomic pairs occurred at much higher temperatures in the unitary regime than previously thought. —JS

Science, this issue p. 452

SOFT ROBOTS

Bilayer microbot powered by humidity

Many plant seeds, including pine cones, open or close in response to changes in humidity. Inspired by these hygroexpansive seeds, Shin *et al.* fabricated a "hygrobot." The hygrobot was composed of a film that was responsive to environmental humidity attached to a film that was not. Aligned nanofibers in the responsive layer swelled when humidity increased, causing the bilayer structure to bend. Decreasing humidity shrunk the moisture-responsive layer, straightening out the robot. Legs with asymmetric friction coefficients converted the bending into directional motion on a moist surface. —RLK

Sci. Robot. **3**, eaar2629 (2018).

ATMOSPHERIC PHYSICS

Up with ultrafine aerosol particles

Ultrafine aerosol particles (smaller than 50 nanometers in diameter) have been thought to be too small to affect cloud formation. Fan *et al.* show that this is not the case. They studied the effect of urban pollution transported into the otherwise nearly pristine atmosphere of the Amazon. Condensational growth of water droplets around the tiny particles releases latent heat, thereby intensifying atmospheric convection. Thus, anthropogenic ultrafine aerosol particles may exert a more important influence on cloud formation processes than previously believed. —HJS

Science, this issue p. 411

HUMAN GENOMICS

Genetic variants provide a nurturing environment

Genetic variants in parents may affect the fitness of their offspring, even if the child does not carry the allele. This indirect

PHOTOS: (TOP TO BOTTOM) HERSHKOVITZ *ET AL.*; RICHARD WHITCOMBE/SHUTTERSTOCK.COM

Published by AAAS

effect is referred to as "genetic nurture." Kong *et al.* used data from genome-wide association studies of educational attainment to construct polygenic scores for parents that only considered the nontransmitted alleles (see the Perspective by Koellinger and Harden). The findings suggest that genetic nurture is ultimately due to genetic variation in the population and is mediated by the environment that parents create for their children. —LMZ

Science, this issue p. 424; see also p. 386

COMPUTER SCIENCE

Libratus versus humans

Pitting artificial intelligence (AI) against top human players demonstrates just how far AI has come. Brown and Sandholm built a poker-playing AI called Libratus that decisively beat four leading human professionals in the two-player variant of poker called heads-up no-limit Texas hold'em (HUNL). Over nearly 3 weeks, Libratus played 120,000 hands of HUNL against the human professionals, using a three-pronged approach that included precomputing an overall strategy, adapting the strategy to actual gameplay, and learning from its opponent. —JS

Science, this issue p. 418

PALEONTOLOGY

Shaping of human brains and behavior

There is little consensus as to when complex cognitive functions and behaviors appeared during human evolution. Traditional explanations propose that these behaviors evolved in tandem with modern human anatomy by 50,000 years ago. Other models postulate a more gradual appearance. Neubauer *et al.* examined endocasts (the interior of the cranial vault) of a sample of early *Homo sapiens* skulls. Although brain size at 300,000 years ago falls within the range of that of modern humans, more globular brains only emerged around 40,000 years ago. Combined with evidence from ancient DNA that indicates the fixing of genes critical to early brain development at the origin of the *H. sapiens* lineage, these findings support models of a more gradual appearance of behavioral modernity. —MSA

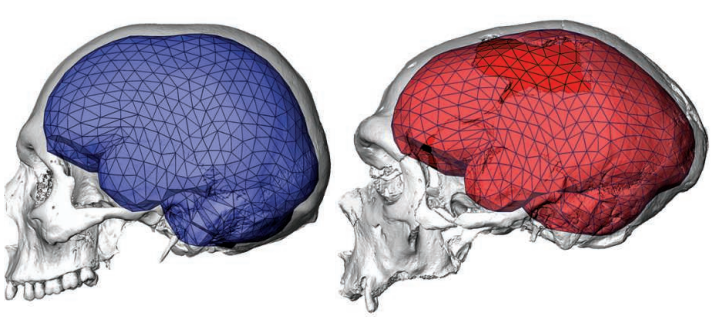
Sci. Adv. 10.1126/sciadv.aao5961 (2018).

HUMAN IMPACTS

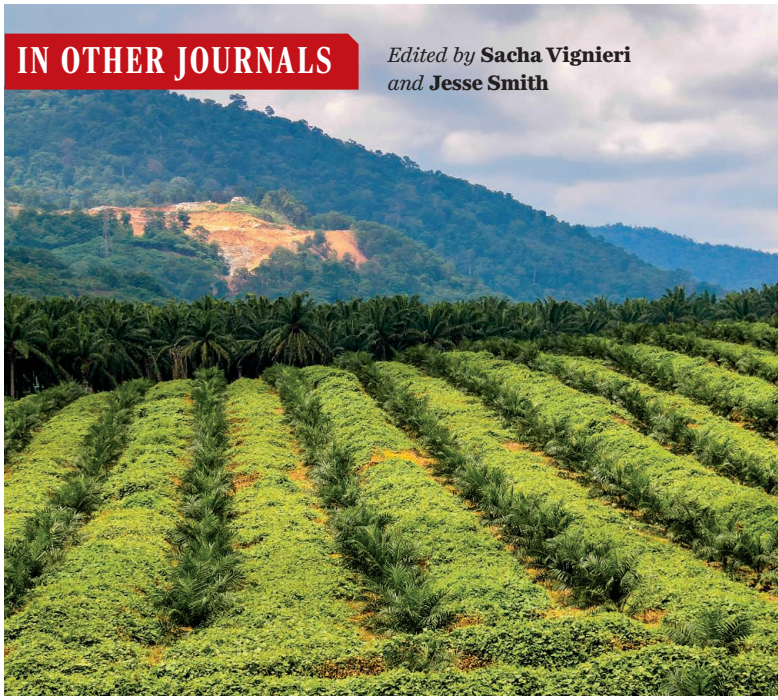
Restrictions on roaming

Until the past century or so, the movement of wild animals was relatively unrestricted, and their travels contributed substantially to ecological processes. As humans have increasingly altered natural habitats, natural animal movements have been restricted. Tucker *et al.* examined GPS locations for more than 50 species. In general, animal movements were shorter in areas with high human impact, likely owing to changed behaviors and physical limitations. Besides affecting the species themselves, such changes could have wider effects by limiting the movement of nutrients and altering ecological interactions. —SNV

Science, this issue p. 466



Brain size and shape changes suggest that modern behaviors developed gradually.



IN OTHER JOURNALS

Edited by **Sacha Vignieri** *and* **Jesse Smith**

PHYSICS

Reviving a quantum spin chain

Quantum systems evolve in time but are expected to experience periodic "revivals"—returns to near-initial state—whose period depends on the energy spectrum of the system. The more complex the spectrum, the longer the period, which makes revivals tricky to observe in systems with long-range interactions. Zeiher *et al.* used a long-range interacting spin chain of 10 ultracold atoms optically "dressed" to a Rydberg state to observe the revivals of magnetization up to very long times. The findings may form the foundation for the use of Rydberg atoms in quantum annealing and similar protocols. —JS

Phys. Rev. X **7**, 041063 (2017).

CHROMOSOME ACTIVATION

Reawakening your X

X chromosome inactivation (XCI) ensures that only one X chromosome in a pair is expressed in a female somatic cell. A heterozygous X-linked mutation therefore affects about half of the somatic cells. In Rett syndrome, such a mutation in methyl-CpG–binding protein 2 (MECP2) causes motor and cognitive abnormalities. Carrette *et al.* identify a combination treatment that "reawakens" the silenced X chromosome, allowing production of normal MECP2 in cultured cells. Treatment involves an antisense oligonucleotide that blocks *Xist*, a noncoding RNA that facilitates XCI. When combined with a chemical inhibitor of DNA methylation (a modification that represses genes), MECP2 expression rose to 5% of the normal level, pointing to the therapeutic potential of partial X chromosome reactivation in the brain. —LDC

Proc. Natl. Acad. Sci. U.S.A. 10.1073/pnas.1715124115 (2018).

GENE REGULATION

Regulation through the core

Enhancers play a pivotal role in gene regulation. These elements can be localized far from the promoter, with looped sequences bringing them together to effect regulated expression. Batut and Gingeras used high-throughput transcription start site mapping and expression patterns to examine promoter activity during early development in five

CREDITS: (LEFT TO RIGHT) NEUBAUER *ET AL.*, *SCI. ADV.* 10.1136/SCIADV.AAO5961 (2018); CALVIN CHAN/SHUTTERSTOCK.COM

TROPICAL DEFORESTATION

Certification of oil palm production

Sustainability certification, whereby commitments are made to produce only sustainable crops, is a tool that is increasingly used to limit the damage to tropical forests caused by large-scale agricultural production. Carlson *et al.* asked how effective this is by assessing the impact of certification schemes for oil palm in Indonesia. A comparison of the performance of certified versus noncertified oil palm plantations over a 15-year period showed that certification reduced the rate of deforestation by 33%—a significant reduction, yet most of the certified areas already retained little residual forest. Certification also failed to reduce fire or peatland clearance. The authors recommend wider adoption of certification schemes in more forested areas to improve the outlook for remaining forest. —AMS

Proc. Natl. Acad. Sci. U.S.A. **115**, 121 (2018).

Certification reduces forest loss, but more needs to be done to make oil palm plantations truly sustainable.

different *Drosophila* species. During early, intermediate, and late embryogenesis, specific core promoters are active, with each core promoter class associated with specific sets of transcription factors. Coarse- and fine-grained developmental regulation are proposed relative to core promoter and enhancer activity, respectively. Genome-wide analysis of long noncoding RNAs (lncRNAs) reveals 3973 promoters, most of which are unannotated, driving expression during embryo development and showing deep conservation of lncRNA promoters. This work shows that enhancers are not the only sequence regulating transcription; core promoters are also key. —BAP

eLife 10.7554/eLife.29005 (2017).

ULTRAFAST OPTICS

Clocking the dynamics of effective mass

Electrons propagating through a solid interact with the crystal lattice, modifying the electronic motion and giving rise to an effective electron mass. The idea of an effective electron mass is based on the assumption of an unbounded crystal lattice, so the question arises of whether this understanding applies at the scale of small electronic devices. Using ultrafast pump-probe spectroscopy, Kasmi *et al.* studied the temporal dynamics of photoemitted electrons from a copper target. They found that the electrons require up to 350 attoseconds to reach their effective mass, equating to a propagation distance of just two atomic layers. The results could have bearing on the performance of shrinking electronic circuits, as well as in correctly interpreting ultrafast photoemission process from solids. —ISO

Optica **4**, 1492 (2017).

ANIMAL MIGRATION

Guiding forces

Although much research has focused on understanding how animals guide their migrations across large regions of the planet, we know very little about this phenomenon, particularly for large, difficult-to-track species. Horton *et al.* tackled this question for some of the largest and farthest-traveling animals on Earth, a set of marine species including humpback whales, white sharks, and northern elephant seals. Using satellite tracking technologies, they found that both individual paths and general routes of travel were often consistent across years, suggesting consistent cues or memories. Further, departure time and routes of travel were correlated with physical cues such as magnetic inclination, moon phase, and gravitational signals. Although purely correlational at this stage, these results suggest that these animals use a suite of planetary indicators to guide them over extremely long distances on consistent routes and with consistent timing. —SNV

Front. Mar. Sci. 10.3389/fmars.2017.00422 (2017).

POLYMER CHEMISTRY

Boron sneaks polarity into polystyrene

The properties of plastics can often benefit from the use of a mix of polar and nonpolar building blocks in their preparation. However, these different building blocks may be chemically incompatible or vary substantially in their solubility and reactivity, complicating synthetic planning. Van de Wouw *et al.* used boron and nitrogen to endow a nonpolar monomer with latent prospects for polarization. The BN aromatic was easily copolymerized with styrene. Peroxide treatment then replaced the dangling boron-bearing ring with a hydroxyl group that would have been challenging to introduce evenly at the outset. —JSY

Angew. Chem. Int. Ed. 10.1002/anie.201711650 (2018).

Planetary cues, such as magnetic inclination and moon phase, guide northern elephant seals and other large vertebrates over long distances and across time.

PHOTO: NNEHRING/ISTOCK.COM

ALSO IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

AGING

Having your longevity and eating too

Although caloric restriction has clear benefits for maximizing health span and life span, it is sufficiently unpleasant that few humans stick to it. Madeo *et al.* review evidence that increased intake of the polyamine spermidine appears to reproduce many of the healthful effects of caloric restriction, and they explain its cellular actions, which include enhancement of autophagy and protein deacetylation. Spermidine is found in foods such as wheat germ, soybeans, nuts, and some fruits and vegetables and produced by the microbiota. Increased uptake of spermidine has protective effects against cancer, metabolic disease, heart disease, and neurodegeneration. —LBR

Science, this issue p. 410

ORGANIC CHEMISTRY

How to get two bonds for the price of one

For more than a century, we have known how to couple aryl chlorides at the sites of their C–Cl bonds to form a single C–C bond. Koga *et al.* found that palladium catalysis can instead activate these C–Cl bonds to attack nearby aromatic C–H bonds in a terphenyl molecular framework. The reaction thereby produces a new ring, fused to the original rings on either side. Polycyclic compounds of this sort are of particular interest in optoelectronics research because of their expansive electron delocalization. —JSY

Science, this issue p. 435

ORGANIC CHEMISTRY

A reaction screen in flowing solvent

Chemists charged with manufacturing pharmaceuticals have recently been exploring the efficiency advantages of continuous flow techniques. Perera *et al.* now show that a flow apparatus can also accelerate reaction optimization earlier in the drug discovery process. They modified a high-performance liquid chromatography system to screen a wide variety of solvent, ligand, and base combinations to optimize carbon-carbon bond formation. Injecting stock solution aliquots of the catalyst and reactants into a carrier solvent stream let the authors vary the main solvent efficiently and scale up the optimal conditions for product isolation. —JSY

Science, this issue p. 429

NANOPHOTONICS

Nanoscale chiral valley-photon interface

Occupation of different valleys within the band structure of some materials can be used to encode information. That information is typically encoded in terms of the chirality or polarization of emitted photons. Gong *et al.* combined a plasmonic silver nanowire with a flake of the transition metal dichalcogenide WS_2 to form a nanophotonic platform for the transfer of solid-state spin into optical information over mesoscopic distances. The direction of light emission from the nanowire was strongly dependent on the spin-orbit coupling of light and the WS_2 layer. Such a highly efficient interface should prove useful for developing valleytronics into a practical on-chip technology. —ISO

Science, this issue p. 443

LIGHT METALS

A framework for more ductile magnesium

Development of ductile magnesium alloys is key to their use in reducing the weight of vehicles and other applications. Wu *et al.* tackle this issue by determining the underlying mechanisms in unprocessed magnesium alloys. Dilute amounts of solutes enhanced certain ductility-improving mechanisms over ones that cause brittle fracture. From this, the authors developed a theory that may be helpful for screening the large number of potential magnesium alloy compositions. —BG

Science, this issue p. 447

QUANTUM INFORMATION

Building an essential quantum component

To build a universal quantum computer—the kind that can handle any computational task you throw at it—an essential early step is to demonstrate the so-called CNOT gate, which acts on two qubits. Zajac *et al.* built an efficient CNOT gate by using electron spin qubits in silicon quantum dots, an implementation that is especially appealing because of its compatibility with existing semiconductor-based electronics (see the Perspective by Schreiber and Bluhm). To showcase the potential, the authors used the gate to create an entangled quantum state called the Bell state. —JS

Science, this issue p. 439; see also p. 393

PALEOANTHROPOLOGY

Earliest modern humans out of Africa

Recent paleoanthropological studies have suggested that modern humans migrated from Africa as early as the beginning of the Late Pleistocene, 120,000 years ago. Hershkovitz *et al.* now suggest that early modern humans were already present outside of Africa more than 55,000 years earlier (see the Perspective by Stringer and Galway-Witham). During excavations of sediments at Mount Carmel, Israel, they found a fossil of a mouth part, a left hemimaxilla, with almost complete dentition. The sediments contain a series of well-defined hearths and a rich stone-based industry, as well as abundant animal remains. Analysis of the human remains, and dating of the site and the fossil itself, indicate a likely age of at least 177,000 years for the fossil—making it the oldest member of the *Homo sapiens* clade found outside Africa. —AMS

Science, this issue p. 456; see also p. 389

NEUROSCIENCE

The neuronal population is the key unit

The responses of pairs of neurons to repeated presentations of the same stimulus are typically correlated, and an identical neuronal population can perform many functions. This suggests that the relevant units of computation are not single neurons but subspaces of the complete population activity. To test this idea, Ni *et al.* measured the relationship between neuronal population activity and performance in monkeys. They investigated attention, which improves perception of attended stimuli, and perceptual learning, which improves perception of well-practiced stimuli. These two processes operate on different time scales and are usually studied using different perceptual tasks. Manipulation of attention and learning in the same behavioral trials and the same neuronal populations revealed the dimensions of population activity that matter most for behavior. —PRS

Science, this issue p. 463

ECOLOGY

Honey bees put pressure on wild pollinators

Honey bees and other pollinators are essential for agriculture. Declining pollinator populations have thus raised alarm among scientists, policy-makers, and

the public. In a Perspective, Geldmann and González-Varo argue that although all pollinators face stressors such as pesticide use, large populations of managed honey bees can themselves put pressure on wild pollinators. It is thus important not to conflate managed honey bees and wild pollinators in conservation efforts. Better honey bee management is particularly crucial during periods when mass-flowering crops are not in bloom, when the bees compete most directly with wild pollinators. —JFU

Science, this issue p. 392

HIV

Indicative integrins in HIV

The gut is thought to be a major viral reservoir in HIV infection. Studies in nonhuman primates have suggested that targeting the α4β7 integrin on T cells may be a viable therapy. Sivro *et al.* now extend these findings to humans by examining HIV acquisition in multiple African cohorts. Higher frequencies of $\alpha4\beta7^+$ circulating T cells before infection were associated with increased HIV acquisition, higher viral load at set point, and more rapid $CD4^+$ T cell decline. Thus, targeting integrins could help reduce the spread of HIV. —LP

Sci. Transl. Med. **10**, eaam6354 (2018).

SYSTEMS IMMUNOLOGY

Committing to cytotoxicity

Cytotoxic $CD4^+$ T cells (CD4-CTLs) were initially identified in patients with chronic viral infections, including dengue virus infection. Patil *et al.* sequenced the T cell receptors of individual $CD4^+$ T cells from human blood to identify precursors that give rise to CD4-CTL cells. CD4-CTL cells underwent marked clonal expansion, and CD4-CTL precursor cells were characterized by high expression of interleukin-7 receptor. These findings should facilitate improved vaccine design in the context of chronic viral infections. —AB

Sci. Immunol. **3**, eaan8664 (2018).

IMMUNOLOGY

A drug to fight two lupus symptoms

The receptor TLR9 is implicated in autoimmune diseases characterized by erroneous recognition of self DNA, such as lupus. Perego *et al.* found that guanabenz, an FDA-approved antihypertensive drug with anti-inflammatory effects, altered cholesterol metabolism so that TLR9 did not reach endosomes, where it is fully activated. Guanabenz treatment reduced symptom severity in a mouse model of lupus. Because many lupus patients also suffer from hypertension, guanabenz and related compounds could have dual benefits. —WW

Sci. Signal. **10**, eaam8104 (2018).

REVIEW SUMMARY

AGING

Spermidine in health and disease

Frank Madeo,*† Tobias Eisenberg,† Federico Pietrocola, Guido Kroemer*

BACKGROUND: As the world population ages, chronic diseases such as diabetes, cardiovascular disease, cancer, and neurodegeneration become ever more prevalent. Interventions that favor healthy aging would constitute powerful strategies with which to limit human diseases that have a broad socioeconomic impact. Fasting regimens such as intermittent fasting or dietary adaptations such as caloric restriction are among the few regimens that extend life and beneficially affect health in all tested model organisms, including rodents and nonhuman primates. However, few people seem capable of changing their dietary routines for extended periods. Thus, supplementation with caloric restriction mimetics (CRMs), which would pharmacologically mimic the beneficial effects of caloric restriction or fasting, has gained attention as an attractive and potentially feasible strategy. The naturally occurring polyamine spermidine, the abundance of which declines during the process of aging, has emerged as a well-tolerable CRM targeting various molecular and physiological age-associated adversities.

ADVANCES: Conceptually, healthy aging requires the retardation of multiple molecular and cellular alterations that drive the aging process and induce age-associated pathologies. These include genomic instability, epigenetic alterations, loss of protein degradation capacity (leading to neurodegeneration), deregulated nutrient sensing, mitochondrial dysfunction, cellular senescence, and chronic inflammation. Spermidine displays pleiotropic effects that include anti-inflammatory properties, antioxidant functions, enhancement of mitochondrial metabolic function and respiration, as well as improved proteostasis and chaperone activity. Many anti-aging effects of spermidine are causally connected to the capacity of this polyamine to induce cytoprotective autophagy. Autophagy ensures general cell homeostasis and proteostasis and is directly involved in the degradation of damaged, potentially toxic organelles and harmful protein aggregates, thus removing and recycling cytoplasmic material that otherwise would accumulate during aging. Consistently, extra supply of spermidine prolongs the life span across species in an autophagy-dependent manner and counteracts age-associated pathologies such as cardiovascular disease, neurodegeneration, and cancer. For instance, dietary spermidine supplementation ameliorates age-induced memory impairment in flies and protects from autoimmune-directed demyelination of neurons in a mouse model for multiple sclerosis. Spermidine also reduces the growth of transplantable tumors, stimulates anticancer immune surveillance in combination with chemotherapy, and suppresses tumorigenesis induced by chemical insults in mice. Moreover, elevated dietary polyamine uptake correlates with reduced cardiovascular and cancer-related mortality in human epidemiological studies. Because spermidine is already present in daily human nutrition, clinical trials aiming at increasing the uptake of this polyamine appear feasible.

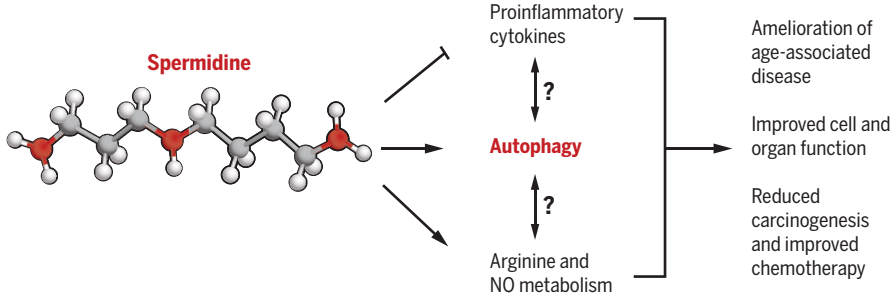
OUTLOOK: Although spermidine induces autophagy and autophagy inhibition curtails many of the health-promoting effects of spermidine, additional mechanisms have been proposed to explain the beneficial effects of spermidine on aging. These potentially autophagy-independent mechanisms include direct antioxidant and metabolic effects on arginine bioavailability and nitric oxide (NO) production. However, it has not been formally determined whether these routes act in a completely autophagy-independent manner or are interrelated with autophagy (in an additive or synergistic way) (see the figure), and it will be important to define actionable molecular targets that explain the beneficial effects of spermidine in diverse pathophysiological settings. In this sense, it will also be of interest to explore synergisms of spermidine with other CRMs that initially act through different mechanisms.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aan2788

Another unresolved enigma resides in the tissue specificity of spermidine-induced health effects. For instance, the mechanisms through which oral spermidine intake can mediate systemic effects on blood metabolites and proteins remain to be elucidated. Similarly, it remains elusive whether spermidine acts exclusively on leukocytes to suppress chronic low-grade inflammation, and to what degree other organs may explain the increased bioavailability of arginine upon spermidine supplementation.

One strong argument in favor of the exploration of spermidine as a CRM in clinical trials is its low toxicity yet strong efficacy. Spermidine is an abundant natural polyamine contained in all organisms from bacteria to men and is naturally present in reasonable but varying amounts in human diets. Nevertheless, based on preclinical results, possible contraindications of spermidine administration such as advanced cancer and renal failure have to be defined. Last, it may be interesting to explore the development of artificial spermidine analogs with increased potency or ameliorated pharmacokinetic characteristics. ■



Schematic outline of the mechanisms of spermidine-mediated health effects. The natural polyamine spermidine has prominent cardioprotective and neuroprotective effects, ameliorates aging-associated metabolic decline, and stimulates anticancer immunosurveillance in animal models. Autophagy is required for several of these health-promoting effects of spermidine. Spermidine also suppresses proinflammatory cytokines and improves the bioavailability of arginine required for NO biosynthesis. It remains an open question whether all these effects depend on the autophagy-stimulatory properties of spermidine.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: frank.madeo@uni-graz.at (F.M.); kroemer@orange.fr (G.K.)
†These authors contributed equally to this work.
Cite this article as F. Madeo *et al.*, *Science* **359**, eaan2788 (2018). DOI: 10.1126/science.aan2788

REVIEW

AGING

Spermidine in health and disease

Frank Madeo,[1,2]*† Tobias Eisenberg,[1,2]† Federico Pietrocola,[3,4,5,6,7] Guido Kroemer[3,4,5,6,7,8,9]*

Interventions that delay aging and protect from age-associated disease are slowly approaching clinical implementation. Such interventions include caloric restriction mimetics, which are defined as agents that mimic the beneficial effects of dietary restriction while limiting its detrimental effects. One such agent, the natural polyamine spermidine, has prominent cardioprotective and neuroprotective effects and stimulates anticancer immunosurveillance in rodent models. Moreover, dietary polyamine uptake correlates with reduced cardiovascular and cancer-related mortality in human epidemiological studies. Spermidine preserves mitochondrial function, exhibits anti-inflammatory properties, and prevents stem cell senescence. Mechanistically, it shares the molecular pathways engaged by other caloric restriction mimetics: It induces protein deacetylation and depends on functional autophagy. Because spermidine is already present in daily human nutrition, clinical trials aiming at increasing the uptake of this polyamine appear feasible.

As the world population ages, chronic maladies such as diabetes, cardiovascular disease, cancer, and neurodegeneration become ever more prevalent because life expectancy is increasing at a quicker pace than does health span (*1*). Interventions that favor healthy aging could constitute powerful strategies to limit human diseases that have a broad socioeconomic impact. Caloric restriction (CR), which is defined as the chronic reduction of calorie intake without malnutrition, is among the few regimens that extend life and beneficially affect health in all tested model organisms, including rodents and nonhuman primates (*2*, *3*). However, it is difficult to set the optimal intensity of CR to avoid undernourishment and other unwanted effects, for which reasons CR is contraindicated in elderly and diseased persons (*1*, *4*, *5*). Moreover, only few people seem capable of changing their dietary routines for extended periods (*6*). Thus, the supplementation of caloric restriction mimetics (CRMs), which would pharmacologically mimic the beneficial effects of caloric or dietary restriction, has gained attention as an attractive and potentially feasible strategy (*7*).

Conceptually, healthy aging requires the attenuation or retardation of several molecular and cellular alterations that drive the aging process and induce age-associated pathologies. These include genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, deregulated nutrient sensing, mitochondrial dysfunction, cellular senescence, stem cell exhaustion, chronic inflammation, and altered intercellular communication (*2*, *8*–*10*). Accordingly, regimens that extend life span of model organisms often display pleiotropic effects that include but are not limited to anti-inflammatory properties, enhancement of mitochondrial metabolic function and respiration, as well as improved proteostasis. This holds true also for the application of CRMs, most of which enhance cytoprotective autophagy (examples of CRMs are summarized in table S1), although additional effects such as on energy metabolism cannot be excluded (*11*). Autophagy ensures general cell homeostasis and proteostasis and is directly involved in the degradation of damaged, potentially toxic organelles or long-lived protein aggregates (*12*). Through controlled sequestration of cytoplasmic material into double-membraned autophagosomes, autophagy directs macromolecules (including proteins, lipids, and nucleic acids) or whole organelles to lysosomal degradation. This process detoxifies and recycles potentially harmful material that accumulates during aging. Consistent with its protective function against aging and disease (*13*–*15*), autophagy is also required for the life-span–extending effects of CR and of several CRMs (*2*, *12*). These CRMs mostly induce autophagy via a common mechanism that involves the deacetylation of cytosolic as well as nuclear proteins (*7*, *16*–*19*). However, the molecular targets accounting for these effects vary among CRMs. CRMs can act as inhibitors of the synthesis of acetyl coenzyme A (CoA), which is essential for enzymatically catalyzed acetylation reactions, for direct inhibitors of acetyltransferases, or as activators of deacetylases (in particular sirtuins) (table S1).

The diamine putrescine and the polyamines spermidine and spermine are ubiquitously occurring polycations that are associated with several important cellular functions and help maintain general cell homeostasis. Polyamines are essential for cell growth and proliferation and tissue regeneration. They bind and stabilize DNA and RNA, have antioxidative activities, modulate enzyme functions, and are required for the regulation of translation. These characteristics have been reviewed elsewhere (*20*–*22*). Polyamines have also been characterized in the control of apoptosis (*20*, *22*) and in the context of learning and memory (*23*). Spermidine acts as a natural autophagy-inducer (*24*) and anti-aging compound that shares many beneficial traits with CR and thus can be considered as a CRM. We review the potential health-promoting effects of polyamines on aging and its comorbidities. Such effects include cardiovascular, neuroprotective, and anti-tumorigenic effects and are induced by dietary or otherwise externally applied spermidine, which will be the primary focus of this article.

Spermidine concentrations decline with age: Bioavailability and metabolism

Polyamine concentrations in mammals are determined by their nutritional supply, synthesis by the intestinal microbiota, uptake, cellular biosynthesis, catabolism, and urinary excretion. Tissue spermidine concentrations decline with age in model organisms as well as in humans (*24*–*27*). This may, at least in part, result from a decline in the biosynthetic activities of polyamine-producing enzymes [reviewed in (*28*)].

Polyamine metabolism and transport

The intracellular spermidine content is the final result of polyamine uptake from the extracellular space, endogenous biosynthesis, catabolism, and excretion (outlined in Fig. 1). Biosynthesis is achieved from the precursor ornithine forming putrescine, spermidine, and spermine (Fig. 1) in tightly regulated step-wise reactions (*20*, *22*, *29*). Catabolism of polyamines involves, on the one hand, the oxidative degradation of spermine to spermidine. On the other hand, the degradation and secretion of both spermidine and spermine requires their acetyl-CoA–dependent acetylation by spermine-/spermidine-N1-acetyltransferase 1 (SSAT1) and subsequent oxidation (Fig. 1) (*30*, *31*). In a recent study, histone deacetylase 10 (HDAC10), a key regulator of autophagy and cell survival (*32*), has been proposed as yet another mediator of polyamines metabolism owing to its ability to deacetylate spermidine (*33*).

Polyamines also interconnect with disease-relevant amino acid metabolism. Synthesis of spermidine and spermine requires the formation of decarboxylated S-adenosyl methionine (dcSAM) from SAM. SAM serves as an important cofactor for methylation of proteins, including histones, as well as of DNA, both of which are necessary for the epigenetic control of gene regulation (Fig. 1). Through the putrescine precursor ornithine, polyamine biosynthesis affects

[1]Institute of Molecular Biosciences, NAWI Graz, University of Graz, Graz, Austria. [2]BioTechMed Graz, Graz, Austria. [3]Equipe 11 Labellisée Ligue Contre le Cancer, Centre de Recherche des Cordeliers, Paris, France. [4]Cell Biology and Metabolomics Platforms, Gustave Roussy Comprehensive Cancer Center, Villejuif, France. [5]INSERM, U1138, Paris, France. [6]University Paris Descartes, Sorbonne Paris Cité, Paris, France. [7]University Pierre and Marie Curie, Paris, France. [8]Pôle de Biologie, Hôpital Européen Georges Pompidou, Paris, France. [9]Department of Women's and Children's Health, Karolinska University Hospital, Karolinska Institute, Stockholm, Sweden.
***Corresponding author. Email: frank.madeo@uni-graz.at (F.M.); kroemer@orange.fr (G.K.)**
†These authors contributed equally to this work.

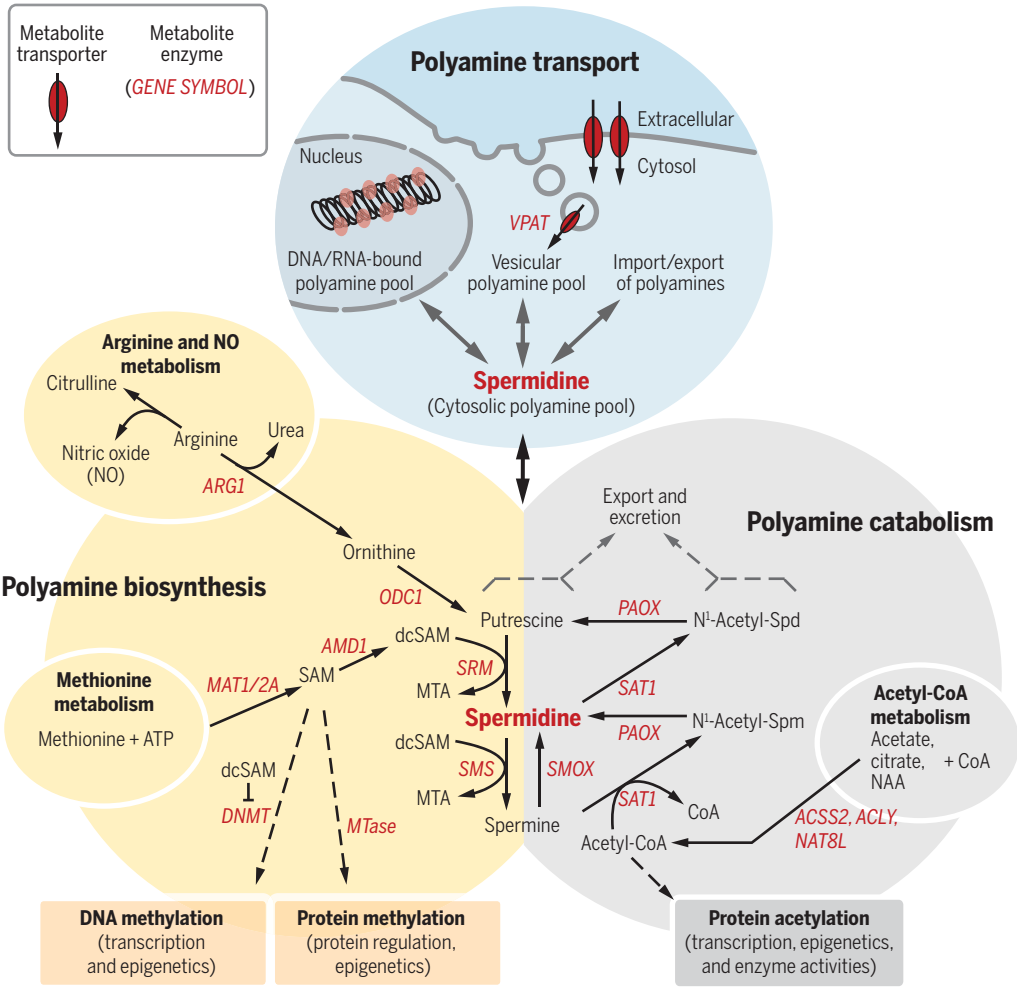


Fig. 1. Regulation of the intracellular spermidine pool. Major routes and enzymes of mammalian polyamine metabolism. The cytosolic spermidine pool results from uptake, biosynthesis, catabolism, and transport. Spermidine is formed from its precursor putrescine or by degradation from spermine. Polyamine biosynthesis connects to arginine and NO metabolism via ornithine (as part of the urea cycle). Spermidine catabolism is mainly mediated through acetylation and subsequent oxidation reactions. Through the cofactors dcSAM (polyamine biosynthesis) and acetyl-CoA (polyamine catabolism), polyamine metabolism interrelates to protein/DNA methylation and protein acetylation, respectively, and thus indirectly influences epigenetic regulation of gene expression. ACLY, ATP-citrate synthase; ACSS2, acetyl-coenzyme A synthetase, cytoplasmic; AMD1, *S*-adenosylmethionine decarboxylase proenzyme; ARG1, arginase-1; ATP, adenosine triphosphate; CoA, coenzyme A; dcSAM, decarboxylated *S*-adenosylmethionine; MAT1/2A, *S*-adenosylmethionine synthase isoform type 1/2; NAA, *N*-acetylaspartate; NAT8L, *N*-acetylaspartate synthetase; NO, nitric oxide; ODC1, ornithine decarboxylase; PAOX, peroxisomal *N*(1)-acetyl-spermine/spermidine oxidase; SAM, *S*-adenosylmethionine; SMO, spermine oxidase; SAT1, spermidine/spermine *N*(1)-acetyltransferase 1; SMS, spermine synthase; Spd, spermidine; Spm, spermine SRM, spermidine synthase; VPAT, vesicular polyamine transporter.

the bioavailability of arginine, which is important for the production of nitric oxide (NO), an important signaling molecule that mediates vasodilation, protects from maladaptive cardiac remodeling, affects mitochondrial biogenesis and function, and has vast immunomodulatory effects (*34*).

The transport of polyamines in mammals is less well understood but may involve plasma membrane transporters as present in yeast and bacteria (*35*). Cellular polyamine uptake and secretion may also be mediated via endocytosis and exocytosis, respectively (Fig. 1). In support of this notion, a vesicular polyamine transporter (VPAT) was identified in astrocytes (*36*) and mast cells (*37*) that may have important neuro- and immune-modulatory implications (*37*, *38*).

Polyamine uptake and excretion

In addition to cellular biosynthesis, two other sources are equally important for the systemic availability of spermidine. These sources are (i) external (oral) uptake with the food and (ii) production by intestinal microorganisms (Fig. 2). Selected unprocessed plant-derived food items are naturally enriched in polyamines, as exemplified by the Durian fruit. Moreover, fermentation processes involving bacteria and fungi used in the food industry cause microbial generation of polyamines, which may contribute to the sometimes subjectively malodorous properties of milk and soy products such as mature cheese and natto, respectively. Ingested spermine and spermidine are quickly adsorbed from the intestinal gut and distributed without degradation (*39*). Thus, diet influences the blood polyamine concentration, which is highly diverse in humans (*40*). The average daily nutritional intake of spermidine varies from ~7 to 25 mg and more, with the highest amounts in the Mediterranean diet, which is often described to improve health in humans (*41*, *42*). These values still lie below the amount of spermidine, when calculated based on the optimal food composition proposed by the Swedish Nutrition Recommendations Objectified (*43*).

In mice, oral supplementation of spermidine increases amounts in whole blood (*44*), serum (*45*), and tissues (*44*). Supplementation of the polyamine precursor arginine (alone or in combination with probiotics) may also suffice to increase spermidine concentrations in the colon and blood (*46*, *47*). Indeed, the intestinal luminal concentration of spermidine critically depends on colonic microbiota in mice (*48*), and this may hold true for humans as well (*49*). Together, these studies outline the importance of dietary polyamines and polyamine-producing bacteria for the bioavailability of spermidine (Fig. 2). Given the high content of polyamines in certain types of food, it appears plausible that a polyamine-rich diet could overcome the age-associated decline of polyamines. Indeed, daily intake of 50 to 100 g of natto over a 2-month period significantly increased the whole-blood spermine content of healthy human volunteers (*40*). A cross-sectional observation in central Italy revealed that whole-blood spermidine and spermine content were decreased in elderly humans but remained at the levels of younger (middle-aged) individuals in healthy nonagenarians and centenarians (*25*). Further studies are warranted to address whether dietary or genetic factors explain the high

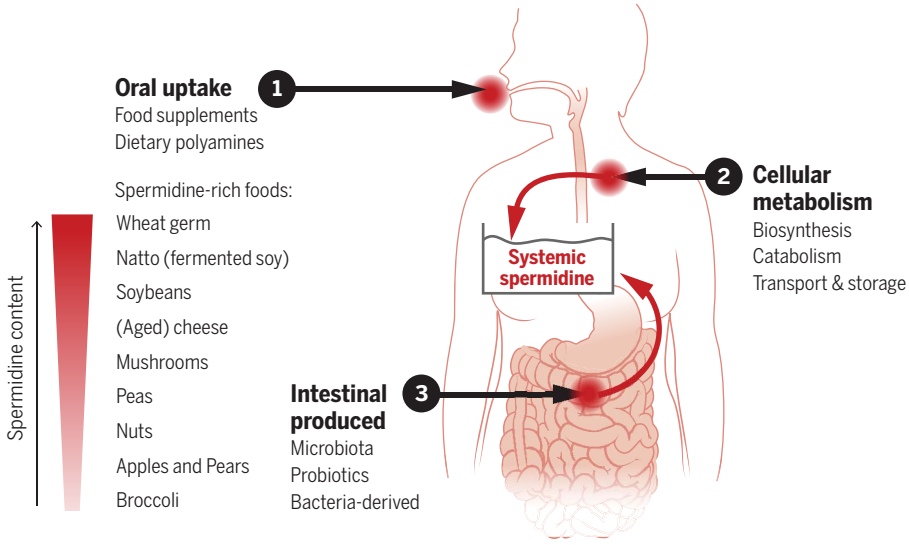


Fig. 2. Sources of systemic spermidine. Scheme depicting the sources crucial for spermidine bioavailability in the whole organism. In addition to cellular metabolism (outlined in Fig. 1), spermidine is taken up orally from dietary sources or produced by commensal gut bacteria. Subsequently, spermidine can be resorbed by intestinal epithelial cells and is distributed through systemic circulation. Examples of spermidine-rich foods are enumerated. Food supplements, including the polyamine precursor arginine and probiotics (polyamine-producing bacteria), increase the intestinal production of polyamines.

spermidine content in healthy centenarians and whether they represent a cause or a consequence of healthy aging.

Systemic concentrations of polyamines are also controlled through renal secretion. Thus, putrescine and acrolein—a toxic product and by-product, respectively, of spermidine and spermine catabolism by serum amine oxidase—increase in the plasma of patients with chronic renal failure (*50*). Therefore, dietary supplementation of spermidine in such patients should be carefully reviewed for potential adverse effects.

Health effects of dietary spermidine

Beyond their capacity to improve cellular fitness—as documented in yeast, plants, and in cultured mammalian cells (table S2)—polyamines (in particular, spermidine) increase life span of multicellular organisms, including nematodes, flies, and mice. Cardiovascular disease and cancer are the two major causes of death in humans and mice. Similarly to CR regimens, which partially protect against these pathologies, spermidine can delay their manifestation and reduce their severity in rodent models.

Life-span extension by spermidine

Health- and life-span–promoting effects are documented for dietary or otherwise externally supplied spermidine (Fig. 3A). Spermidine supplementation confers life-span extension both to invertebrate model organisms (*19*, *24*) and to mice (*44*, *51*). Moreover, a diet rich in polyamines reduces mortality of aged mice (*52*). As true for many other CRMs, spermidine extends life span in a sex-independent manner (table S1). Supplementation of polyamine-producing probiotic bifidobacteria—alone or, moreso, in combination with the polyamine precursor arginine—increases blood concentrations of spermidine and spermine and decreases mortality and the prevalence of cutaneous pathologies in old mice (*46*, *47*). Furthermore, oral spermidine administration counteracts the age-associated disruption of circadian rhythm in mice (*45*). Spermidine may also ameliorate menopausal decline. Indeed, bone loss induced by ovariectomy in mice, a model of postmenopausal osteoporosis, is prevented with oral spermidine supplementation via inhibition of the formation of (bone-resorbing) osteoclasts (*53*). The total amount of food polyamines significantly correlates with human life expectancies across distinct Asian countries (*54*), although this epidemiological study did not adjust for confounding factors typically associated with longevity. Future well-controlled epidemiological studies are warranted to explore the effects of dietary spermidine on human health span and life span.

Polyamine effects on tumorigenesis

Polyamines are essential for cell proliferation and growth, and dysregulation of polyamine metabolism is a defining signature of many tumor types; increased polyamine concentrations caused by enhanced biosynthesis can be found in skin, breast, colon, lung, and prostate cancers (*35*). Given the proliferation-enhancing and cytoprotective effects of polyamines on cultured human cancer cells or xenografted human tumors evolving in immunodeficient mice [reviewed in (*55*)], polyamines might have procarcinogenic properties. Efforts have been made to suppress polyamine biosynthesis by inhibiting ornithine decarboxylase (ODC) with difluoromethylornithine (DFMO) for the treatment of established cancers in mice, but clinical trials using DFMO have been abandoned because of its toxicity (*35*). Potentially less toxic, a competitive inhibitor of AdoMetDC (SAM486A), leading to low spermidine concentrations, has been tested in clinical trials for various tumor types, with moderate success in non-Hodgkin's lymphoma (*35*, *56*).

In contrast to potential procarcinogenic properties of polyamines, spermidine supplementation can reduce tumorigenesis in mice. Oral supplementation of probiotic polyamine-producing bacteria (*Bifidobacterium animalis* subsp. *Lactis* LKM512) lowers the incidence of visible skin tumors in aging Crj:CD-1 female mice (*47*). Dietary spermidine reduces the severity of liver fibrosis and the incidence of hepatocellular carcinomas induced by chemical insults in mice (*51*). Spermidine administration also slows the growth of CT26 colorectal tumors transplanted into immunocompetent mice (*57*). Increased intake of dietary polyamines (using chow enriched in spermidine, spermine, and putrescine) causes a delay in chemically induced tumorigenesis in young BALB/c male mice, although the maximum size of tumors increased (*58*). Thus, in mice polyamines may inhibit colon carcinogenesis yet favor tumor growth, once a cancer has developed. In humans, one study suggested that dietary intake of polyamines above the median could be associated with an elevated risk of developing colorectal adenoma, particularly in women (*59*). Nevertheless, a subsequent prospective study performed by the same authors failed to confirm this positive association and rather revealed an inverse correlation of high polyamine intake with the risk of colorectal cancers, at least in overweight women with a body mass index (BMI) ≤ 25 (*60*). Interventional studies are needed to explore the possible effect of dietary spermidine on cancer risk.

Spermidine supplementation reduces the growth of transplantable tumors in mice treated with chemotherapies. This spermidine effect is shared by other CRMs (*61*) as well as by fasting or hypocaloric diets (*62*, *63*) and is mediated by the stimulation of immunosurveillance. Thus, spermidine and other CRMs enhance the anticancer immune response as they deplete immunosuppressive cells such as regulatory T lymphocytes (T_{reg} cells) from the tumor bed (*61*). Tumors growing in mice that lack cytotoxic T lymphocytes fail to reduce their growth in response to spermidine or other CRMs (*61*). Beyond its effects on adaptive immunity, spermidine synthase (SRM)–mediated spermidine production appears to be a determinant for antitumor actions of tumor-associated macrophages (TAMs) in the context of colorectal cancer progression (*57*). These findings may explain why external supply of spermidine has chemopreventive effects in vivo (that are likely mediated by immunostimulatory effects), although it enhances the proliferation of cancer cells in vitro.

Future studies should address associations of dietary spermidine with other types of cancer, especially in light of the potential requirement of polyamines for cancer cell growth. Whether a

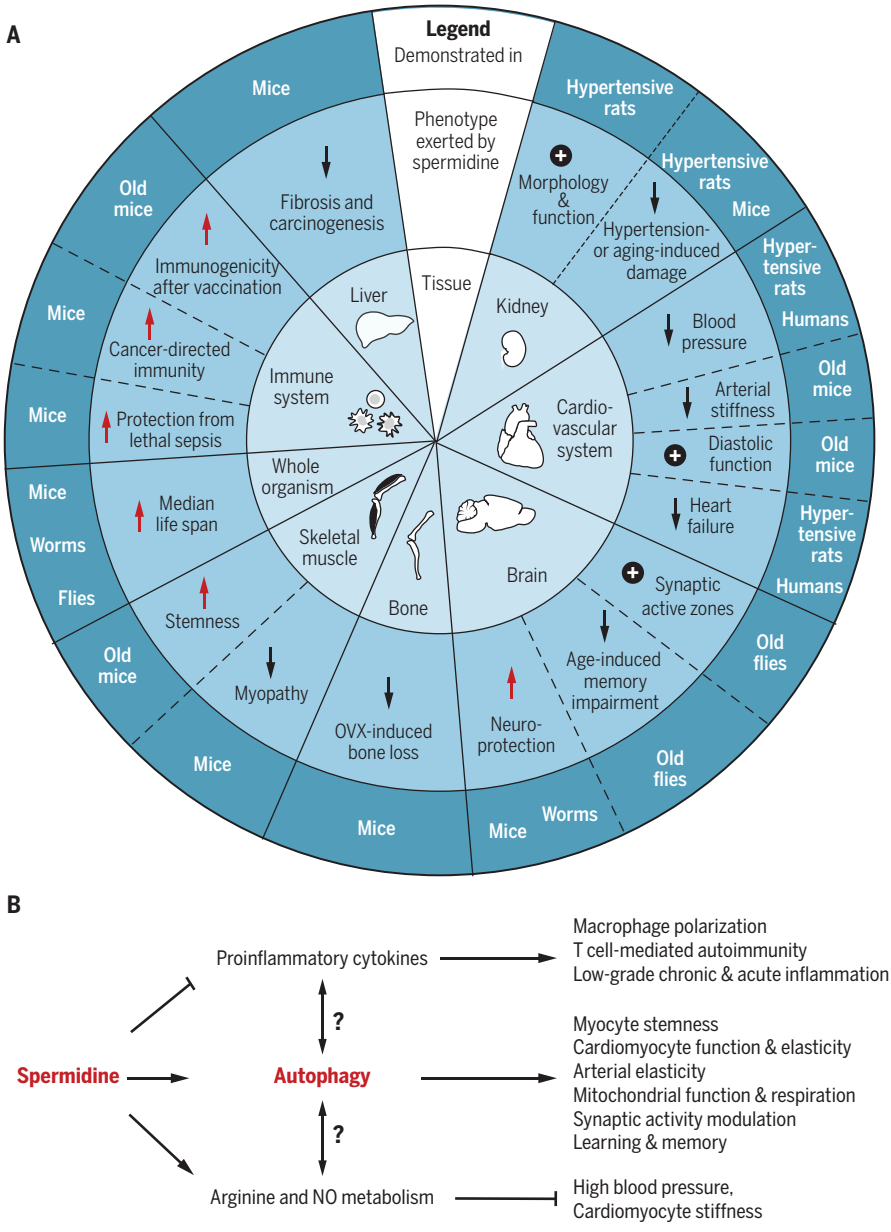


Fig. 3. Spermidine-mediated health effects. (**A**) Summary of effects elicited by dietary or otherwise supplemented spermidine in different organ systems. Plus symbols indicate "improved." (**B**) Spermidine-enhanced autophagy is required for several of the health-promoting effects presented in (A). Spermidine also suppresses proinflammatory cytokines and improves the bioavailability of arginine required for NO biosynthesis, mediating immunomodulatory and antihypertensive effects. It remains an open question to which extent autophagy contributes to spermidine-mediated changes in cytokine production and arginine or NO metabolism.

diet low in polyamines slows growth of specific cancers and whether a polyamine- or specifically spermidine-rich diet would accelerate tumor growth in human patients remain open questions, requiring in-depth investigation before such diets can be broadly recommended, at least for this specific risk group.

Spermidine in cardiovascular and muscle-related disease

Dietary spermidine protects from cardiac aging. It improves diastolic function, left ventricular elasticity, and mitochondrial function in old mice (*44*). First results from epidemiological studies corroborate these findings in humans: Intake of dietary spermidine (or that of spermidine and spermine combined) inversely correlates with the incidence of cardiovascular disease (CVD) and death in the Bruneck cohort (*44*). A cross-sectional regression meta-analysis of nutritional polyamine content with CVD-caused mortality rates by using publicly available data from 48 Western countries identified negative associations of spermidine and spermine with CVD (*64*).

Spermidine reversed age-induced arterial stiffness with a reduction in oxidative damage of endothelial cells in old mice (*65*) and alleviated the formation of atherosclerotic plaques in apolipoprotein E–deficient ($ApoE^{-/-}$) mice fed a high-fat diet (HFD) for 20 weeks (*66*). In Dahl salt-sensitive rats fed a high-salt diet (a model of hypertensive heart failure), oral supplementation of spermidine reduced high blood pressure and delayed the transition to heart failure (*44*), further documenting the antihypertensive (*67*) and vascular health–promoting (*65*, *66*) functions of dietary spermidine. Humans belonging to the higher tertile of dietary spermidine uptake in the Bruneck cohort had lower diastolic and systolic blood pressure as compared with those of the lower tertile subjects (*44*).

One of the hallmarks of aging is the decline in stem cell function, resulting in impaired tissue regeneration and immunosenescence (*8*). As a first indication that polyamines might facilitate the manifestation of stem cell–like features, spermidine favored the reprogramming of somatic cells to induced pluripotent stem cells (iPSCs) of mouse embryonic fibroblasts in vitro (*68*) and promoted hair growth and epithelial stem cell function in human hair follicle organ cultures (Fig. 4F) (*69*). Short-term administration of spermidine to old mice reversed the age-associated defect of autophagy and mitophagy in muscle stem cells (satellite cells), preventing their senescence and improving muscle regeneration (*70*). Similarly, spermidine reactivated autophagy and ameliorated the myopathic defects of collagen-VI–deficient mice, which are normally characterized by invalidating mitochondrial defects (*71*). Skeletal muscle atrophy in D-galactose–treated rats was inhibited by spermidine application (or spermidine combined with exercise), concomitant with induction of autophagy and mitochondrial improvements (*72*). The spermidine-mediated ultrastructural and functional improvement of mitochondria from aged cardiac muscles (*44*) and from skeletal muscle (stem) cells (*70*–*72*) further support the potential utility of spermidine in the treatment of muscle-related disorders.

Polyamines in metabolic syndromes

Whether polyamines might be useful in treatment of obesity and type 2 diabetes is an important topic for future research. Daily administration of spermine to mice eating a HFD prevented adiposity and improved glucose tolerance (*73*). The white adipose tissue (WAT) or liver-specific depletion of nicotinamide *N*-methyltransferase (Nnmt) [an enzyme that methylates the nicotinamide adenine dinucleotide (oxidized form) (NAD^+) precursor nicotinamide by using SAM as methyl-group donor and therefore links polyamines metabolism to NAD^+ metabolism and Sirtuin signaling (Fig. 1)] rendered mice resistant against HFD-induced obesity (*74*); the down-regulation of Nnmt resulted in an augmented polyamine flux (as corroborated by the simultaneous activation of ODC-mediated spermidine biosynthesis and SSAT1-driven spermidine elimination) and led to an increased energy expenditure

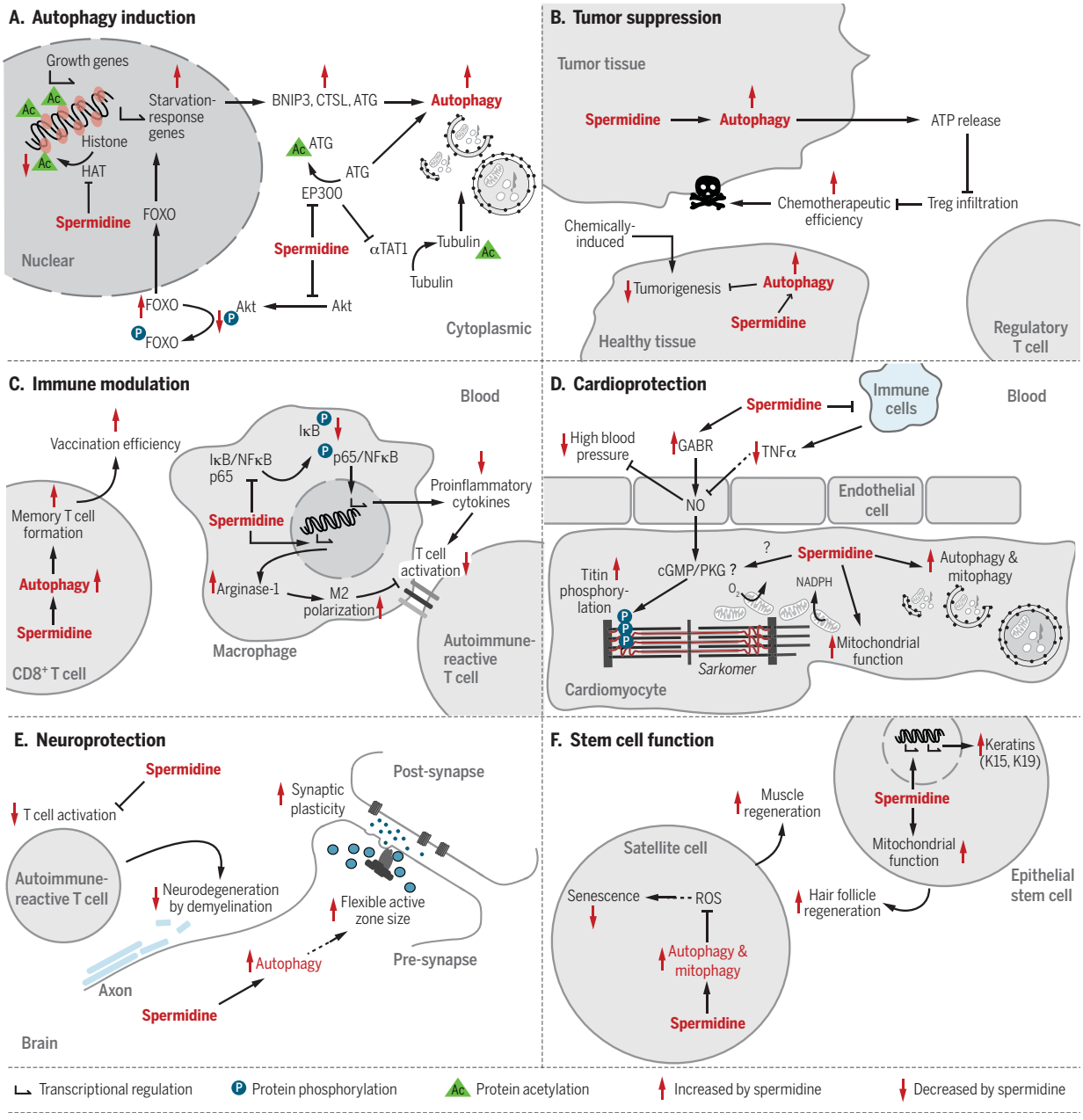


Fig. 4. Cellular and molecular mechanisms of spermidine-mediated health protection. Mechanistic models of the cellular and molecular effects elicited by spermidine through transcriptional, posttranslational (affecting protein acetylation and phosphorylation), as well as metabolic effects. (**A**) Rapid autophagy induction by spermidine administration through inhibition of the acetyl transferase EP300, primarily resulting in autophagy-relevant cytosolic protein deacetylation. Sustained autophagic control by spermidine is mediated through induction of autophagy-relevant gene transcription. This involves the regulation of FOXO transcription factor as well as inhibition of histone acetyl transferases, resulting in epigenetic transcriptional reprogramming. (**B**) Spermidine may suppress tumorigenesis through induction of autophagy in healthy cells. In autophagy-competent tumor cells, spermidine favors the autophagy-dependent release of ATP, which in turn favors immunosurveillance. (**C**) Anti-inflammatory effects of spermidine are explained through its effects on macrophages, promoting M2 polarization and the suppression of NFκB-dependent proinflammatory cytokines. These inhibitory macrophages then suppress autoimmune-reactive T cells. At the same time, spermidine favors formation of $CD8^+$ memory T cells via induction of autophagy. (**D**) The suppression of circulatory cytokines, such as TNFα, also contributes to cardiovascular protection, possibly via a concerted action with arginine-derived nitric oxide, which leads to vasodilation and promotes the cGMP/cGMP-dependent protein kinase (PKG)–dependent phosphorylation status of titin. Spermidine-enhanced autophagy and mitophagy also contribute to cardiomyocyte elasticity and mitochondrial functionality. (**E**) The inhibitory effect of spermidine on autoimmune-reactive T cells (C) further translates into prevention of neurodegeneration by demyelination. Neuroprotection is also mediated via autophagy-dependent proteostasis of presynaptic active zones, assuring the maintenance of synaptic plasticity. (**F**) Suppression of stem cell senescence by spermidine depends on autophagy (in satellite cells), whereas spermidine promotes mitochondrial function and production of keratins in epithelial stem cells. These stemness-enhancing effects ensure muscle and hair follicle regeneration, respectively. Red arrows (up or down) indicate changes (increased or decreased, respectively) that are observed after spermidine supplementation.

(74). Similarly, whole-body depletion of the RNA polymerase III repressor MAF1 (which leads to a reduced expression of Nnmt) increases life span and confers resistance to HFD-induced obesity associated with enhanced polyamine flux and activation of autophagy (75). Of note, autophagy is required for full weight loss upon acute starvation and counteracts weight gain– and obesity-related pathologies in mice fed hypercaloric diets. Spermidine attenuated weight gain and the comorbidities of obesity induced by hypercaloric regimens, correlating with autophagy induction in WAT (76).

Neuroprotection by spermidine

Autophagy (and particularly mitophagy) have a major role in maintaining normal brain function and protecting from neurodegeneration (13, 77), and spermidine supplementation exerts neuroprotective effects in vivo (Fig. 3A). In flies, spermidine feeding protects from age-induced memory impairment (26) and loss of locomotor activity (78) in an autophagy-dependent manner. This has been explained by the autophagy-dependent rejuvenation of synaptic active zone composition (Fig. 4E), which is important for synaptic flexibility and plasticity (79). In a mouse model of experimental autoimmune encephalomyelitis, a model for multiple sclerosis, oral supplementation of spermidine attenuates disease progression and improves visual functions through the reduced demyelination of optic nerve and spinal cord and decreased loss of retinal ganglion cells (80, 81). Similarly, spermidine promotes optic nerve regeneration and retinal ganglion cell survival after optic nerve injury in vivo in mice (82) and blunts retinal degeneration in a mouse model of normal tension glaucoma (83). Spermidine biosynthesis induced through up-regulation of arginase-1 is required for spontaneous axonal regeneration after peripheral lesion of dorsal root ganglion neurons in mice (84). Spermidine further prevents α-synuclein neurotoxicity in invertebrate models (85), reduces frontotemporal lobar dementia (FTLD-U) in mice expressing transgenic transactive response DNA–binding protein 43 kDa (TDP-43) (86), and attenuates laurate-induced cerebral small vessel disease (87). Striatal injection of spermine improved the recognition memory deficit in a rodent model of Huntington's disease (88), and up-regulation of colonic polyamines through orally administered arginine in combination with the probiotic bifidobacteria LKM512 improved the spatial learning and memory capabilities of old mice (46).

Altogether, these studies suggest a wide range of neuroprotective effects of exogenously applied spermidine with relevance to several neurodegenerative motor disorders and dementias. The underlying mechanisms are discussed below in detail.

Cellular and molecular mode of action of spermidine

Spermidine supplementation induces autophagy and exerts additional and potentially autophagy-independent metabolic and transcriptional effects in vivo. These changes may explain the disease-modulatory and anti-aging mode of action of polyamines.

Autophagy and mitophagy induction through spermidine administration

Spermidine administration stimulates autophagy in aging yeast, flies, and worms (24) and in cultured mammalian cells (18, 19, 70). In mice, acute application of spermidine (for example, by means of intraperitoneal injection) induces autophagy in vivo in multiple tissues such as heart, liver, and muscle within a few hours (4 to 24 hours) (19, 61). Concomitantly, spermidine induces major changes in the metabolome of the plasma, heart, skeletal muscle, and liver that resemble those triggered by other CRMs or fasting (61), which is in line with a common mechanism of action of these pharmacological and nutritional interventions. Oral supplementation of spermidine in drinking water triggers autophagy after 2 to 4 weeks in cardiac tissue (44) and reverses the age-dependent decline of autophagy in the aorta (other tissues remain to be tested) (65). At least in cardiomyocytes in vivo and in various cultured cell lines, spermidine also increases the autophagy-dependent selective degradation of mitochondria (that is, mitophagy) and thus contributes to mitochondrial health and functionality (44, 70, 72, 89).

Given the predominantly cytoprotective role of autophagy and mitophagy, the induction of autophagy may explain the health-promoting effects of spermidine (Fig. 3B), like those of other CRMs (7, 90). In support of this, genetic inhibition of autophagy abolishes spermidine-induced extension of life span in flies, worms (24), and mice (51) and prevents restoration of memory performance and synaptic flexibility in aged flies (26, 79). Inhibition of autophagy partially reduces the spermidine-mediated resistance to paraquat-induced loss of locomotor activity in *Drosophila* (91). Moreover, spermidine fails to promote cardioprotective effects in cardiomyocyte-specific $Atg5^{-/-}$ mice (44) and does not restore stem cell–proliferative function in autophagy-deficient ($Atg7^{-/-}$) Pax7-expressing cells (Pax7 represents a marker and regulator of muscle progenitors and satellite cells) (70). Last, the spermidine-mediated reduction of atherosclerotic lipid accumulation and necrotic core formation that is normally observed in $ApoE^{-/-}$ mice is absent in $ApoE^{-/-}Atg7^{-/-}$–deficient (autophagy-incompetent) animals (66).

Several CRMs induce autophagy by decreasing acetylation of multiple cellular proteins (7). Spermidine diminishes cytosolic protein acetylation by inhibiting the activity of several acetyltransferases, including that of E1A-associated protein p300 (EP300) (19, 24), and EP300 inhibition is sufficient to acutely induce autophagy (18). EP300 directly inhibits acetylation of several autophagy-essential autophagy-related (ATG) proteins (92, 93) and indirectly stimulates deacetylation of tubulin by inhibiting α-tubulin acetyltransferase 1 (αTAT1) (94). Therefore, the inhibition of EP300 by spermidine causes deacetylation of ATG proteins and increases acetylation of tubulin, thus stimulating autophagic flux (Fig. 4A). In addition, autophagy induction and extension of life span in aging mice depends on spermidine-mediated stabilization of the pro-autophagic microtubule-interacting protein microtubule-associated protein 1S (MAP1S) and may involve the depletion of cytosolic HDAC4 (51). Long-term, chronic effects of spermidine on the rate of basal autophagic flux may be additionally explained by transcriptional effects on autophagy-related genes (Fig. 4A). For instance, the transcription factor FoxO3, which is inhibited by protein kinase B (PKB, best known as Akt), is necessary and sufficient for the induction of autophagy in mouse skeletal muscle in vivo (95). The spermidine-mediated reactivation of muscular autophagy in collagen VI–deficient mice coincides with the dephosphorylation (and hence inactivation) of Akt and in turn with increased transcription of FoxO3-dependent target genes (71). Support for direct inhibition of acetyltransferases by spermidine comes from in vitro activity assays, in which nuclear extracts were incubated with spermidine (24) and enzymatic assays with recombinant EP300 protein that reveal spermidine as a competitive inhibitor that is particularly active when acetyl-CoA concentrations are low (18). Spermidine may also reduce the activity of acetyltransferases by reducing the availability of acetyl-CoA because the catabolism of spermidine requires acetylation and hence consumes acetyl-CoA (Fig. 1). Depletion of the cytosolic-nuclear pool of acetyl-CoA induces autophagy in cultured mammalian cells, and in mice, aging yeast, and flies (16, 17, 96), and thus may represent an important mode of action of CRMs that induce protein deacetylation (table S1) (7). Given the pleiotropic action of spermidine on autophagy-related gene transcription, protein stability, and posttranslational regulation, it will be important to molecularly define the key events that are essential for autophagy induction by spermidine.

The protein deacetylase Sirtuin 1 inhibits (97) whereas cytosolic acetyl-CoA and EP300 promote the activity of the autophagy-inhibitory mechanistic target of rapamycin complex 1 (mTORC1) (17). Indeed, spermidine and several other CRMs cause mTORC1 inhibition (7, 17, 18) and activate phosphorylation of 5′ adenosine monophosphate–activated protein kinase (AMPK) (17), which antagonizes mTORC1 at the functional level and may further facilitate autophagic response to these CRMs. Induction of mitophagy by spermidine has also been linked to ataxia-telangiectasia mutated protein kinase (ATM)–dependent activation of the phosphatase and tensin homolog (PTEN)–induced putative kinase 1 (PINK1)/Parkin-signaling pathway (89), which has an important role in mitochondrial quality control (98). Spermidine induces mitochondrial depolarization and mitophagy in cultured human fibroblasts but does not do so in cells lacking ATM or treated with the ATM inhibitor KU55933. However, the precise mechanism by which spermidine favors the activation (and phosphorylation) of AMPK and ATM remain elusive. Spermidine-treated cells exhibit

changes in phosphorylation of multiple proteins (*99*), pointing to a sophisticated cross-talk of the protein acetylation and deacetylation machinery and protein kinase cascades that requires further in-depth investigation.

Immunostimulatory and anti-inflammatory effects of spermidine

The success of current anticancer treatments, including chemotherapies, essentially relies on the immune cell–dependent clearance of tumor cells. This is particularly important for the long-term efficacy of cytotoxic agents that induce immunogenic cell death (*100*–*102*). Autophagy is required for cancer cells to adapt to cell-intrinsic and -extrinsic stress. However, premortem autophagy also promotes immune recognition of cancer cells because it facilitates the release of adenosine 5′-triphosphate (ATP), which serves as a chemotactic factor for myeloid cells, including dendritic cell precursors (*103*, *104*). By reducing the stimulation of immunosuppressive adenosinergic receptors, autophagy also prevents the recruitment of T_{reg} cells into the tumor bed (*105*). Autophagy induction by spermidine or by the CRM hydroxycitrate favors the release of ATP from neoplastic cells in mice and therefore ameliorates immunosurveillance, resulting in a delay in tumorigenesis of autophagy-competent (but not autophagy-deficient) mutant KRAS-induced lung cancers in vivo (Fig. 4B) (*61*). Thus, the administration of spermidine or hydroxycitrate, similar to that of acute nutrient deprivation (*106*), can favor chemotherapeutic efficiency against autophagy-competent cancers. Spermidine also directly improves the generation and function of memory lymphocytes, similar to what is known for other autophagy inducers such as rapamycin (*107*) or metformin (*108*). T cell function declines with age, and this decline is associated with impaired autophagy, which can be observed in cells isolated from elderly human donors (*109*). In old mice, $CD8^+$ T cell responses to influenza vaccination are restored by oral spermidine treatment in an autophagy-dependent manner (*110*), suggesting important roles for autophagy and spermidine in memory cell formation (*111*). Of note, in isolated T cells spermidine apparently induces autophagy independently from mTORC1 because it fails to affect the phosphorylation of the mTORC1 substrate ribosomal S6 kinase (*110*).

Exogenously supplied spermidine and spermine suppress the secretion of proinflammatory cytokines in numerous pathophysiological settings (table S2). These effects are in part secondary to autophagy induction because autophagy has broad anti-inflammatory effects that rely in part on the inhibition of inflammasome activation (*112*). Dietary spermidine suppresses the age-associated chronic low-grade increase in plasma concentrations of tumor necrosis factor–α (TNF-α) and other cytokines in mice (*44*, *46*) and inhibits lethal sepsis (*113*).

Lymphocyte function–associated antigen–1 (LFA-1) is crucial for regulating immune cell adhesion and migration, and its expression increases with age in human blood cells (*114*). Spermidine and spermine both suppress LFA-1 expression in human lymphocytes in vitro and, when orally supplied, in aging Jc1:ICR male mice in vivo (*58*, *115*). Blood spermine concentrations inversely correlate with abundance of LFA-1 in blood obtained from healthy human volunteers (*114*). The underlying mechanism may involve age-associated alterations in DNA methylation (*58*, *115*) or inhibitory effects of polyamines on proinflammatory transcription factors (Fig. 4C), including nuclear factor–κB (NFκB) in macrophages (*81*).

Neuromodulatory and neuroprotective mechanisms of spermidine

The neuromodulatory actions of spermidine and spermine result from their interaction with ionotropic receptors—most importantly, *N*-methyl-D-aspartate (NMDA) receptors [reviewed in (*23*)]. Neuroprotection conferred by spermidine may also result from increased autophagy in neuronal or glial cells, as well as from suppression of maladaptive inflammation. Old flies deficient in Atg7 fail to recover their memory function and synaptic flexibility when fed with spermidine (*26*, *79*). This is consistent with a general age- and neuroprotective function of autophagy in this model organism (*12*, *13*). At first sight counterintuitively, spermidine causes a decrease in presynaptic active zone scaffold proteins (Bruchpilot and Rim-binding protein) and a concomitant reduction in synaptic vesicle release (Fig. 4E) (*79*). Through this effect, spermidine restores synaptic dynamics that are otherwise pushed to their dynamic limits at old age, causing memory impairments. Thus, spermidine appears to ensure proteostasis of synaptic active zones in an autophagy-dependent fashion. In addition, spermidine prevents the toxicity of transgenic α-synuclein expression in flies and nematodes (*85*), which is consistent with the idea that (spermidine-enhanced) autophagy clears polyubiquitin-associated and other neurotoxic protein aggregates.

In spite of the potent neuroprotective capacity of spermidine in invertebrates, it remains elusive whether spermidine is able to cross the mammalian blood-brain barrier or even to regulate its integrity (*116*, *117*). At least after forebrain ischemia, spermidine is transported into several areas of the rat brain (*117*). Alternatively, spermidine may act on nervous tissue via systemic routes that involve regulation of inflammatory cytokines and immune cells capable of invading cerebral tissue. In line with this idea, the preventive action of spermidine against axon demyelination (*80*) is mediated through inhibition of autoimmune-reactive (demyelinating) T cells (*81*). Spermidine prevents activation of autoimmune-reactive T cells indirectly by favoring the polarization of circulatory macrophages to T cell inhibitory M2 cells. This process is mediated by NFκB-dependent suppression of proinflammatory cytokines and by transcriptional activation of arginase-1 (Fig. 4, C and E), an important modulator of both adaptive and innate immune responses (*118*).

In mouse models of Alzheimer's disease, increased monoamine oxidase type B (MAOB) activity has been measured in reactive astrocytes. This MAOB activity favors the conversion of putrescine to the gliotransmitter γ-aminobutyric acid (GABA), which drives Alzheimer's-associated pathology in APP/PS1 mice that carry mutations in the amyloid precursor protein (APP) and the γ-secretase presenilin-1 (PS1) (*119*). MAOB protein levels are elevated in the brain of Alzheimer's patients, and a higher activity is observed in plaque-associated astrocytes (*120*). Therefore, potential adverse effects of dietary polyamines in such patients should be carefully considered and weighed against their neuroprotective potential.

Mode of action of cardiovascular protection by spermidine

Beyond the induction of autophagy and mitophagy that result in improved mitochondrial structure and function in cardiomyocytes (*44*), spermidine protects from cardiovascular pathologies via complementary pathways (Fig. 4D). In Dahl rats fed a salt-rich diet, spermidine increases the bioavailability of arginine (likely because arginine is no more needed for endogenous polyamine biosynthesis), which may translate into enhanced production of the vasodilator NO, reducing arterial hypertension (*44*). At the same time, spermidine ensures cardiomyocyte elasticity through phosphorylation of the myosin filament–associated protein titin (*44*). This molecular event is favored by suppression of proinflammatory cytokines and the availability of nitric oxide (NO) with subsequent guanosine 3′,5′-monophosphate (cGMP)–dependent protein kinase–dependent signaling. The capacity of spermidine to reverse arterial aging has also been linked to enhanced NO bioavailability as well as to the suppression of oxidative damage in endothelial cells (*65*). Thus, metabolic effects of spermidine may include the feedback inhibition of endogenous polyamine biosynthesis and limiting consumption of arginine for ornithine biosynthesis, thus facilitating arginine-dependent NO biosynthesis (Fig. 1). The antioxidant action of spermidine on in vitro–cultured mouse arteries appeared to be autophagy-dependent because it was blocked by the autophagy inhibitor chloroquine (*65*). Thus, it remains to be investigated whether the effects of exogenous spermidine on arginine and NO metabolism and autophagy occur in an independent manner or are somehow interlinked.

Outlook: Diagnostic and therapeutic potential of spermidine in humans

Several pathologies have been associated with increased polyamine concentrations, raising the possibility of considering polyamines as disease biomarkers. This applies in particular to cancer, neurodegenerative disease, stroke, and renal failure [reviewed in (*121*)], as well as heart failure and cardiac infarction (*122*, *123*). Furthermore, disturbed polyamine and arginine metabolism is observed in mouse models of Alzheimer's disease (*124*) and in human patients with mild cognitive impairments (MCIs) (*125*). Several metabolites

including spermidine were differentially affected in stable MCI patients and patients that later developed Alzheimer's disease (*125*). In subjects developing Alzheimer's disease, putrescine may be preferentially channeling into production of spermidine and spermine, whereas the stable MCI condition favors the conversion of putrescine to *N*-acetylputrescine and 4-aminobutanal (*125*). In light of these findings, quantitative high-throughput methods to analyze polyamines and related metabolites (*122*, *126*) could be used on vast sample collections to further evaluate the relationship of such compounds to disease evolution.

Increased concentrations of whole-blood spermidine and spermine are associated with longevity in healthy nonagenarians and centenarians (*25*). Therefore, the sole measurement of the polyamines spermidine and spermine in blood or tissue specimens will likely not suffice as biomarkers for general health or specific diseases. Moreover, the association of increased spermidine with various human pathologies does not necessarily argue in favor of its causal involvement. On the one hand, increased biosynthesis of spermidine could be a homeostatic response to stress that is activated to prevent excessive damage (such as by inducing autophagy). This has been proposed, for instance, for the compensatory cytoprotective response of spermidine observed in uncoupling protein 1 (UCP1) transgenic mice, which is a model for skeletal muscle–specific mitochondrial dysfunction (*127*). On the other hand, increases in polyamine biosynthesis could result from disease-associated alteration of other interlinked metabolic pathways. The complexity of age-associated effects on polyamine metabolism is further emphasized by diverse and region-specific changes in polyamine levels of memory-associated brain structures (*128*).

Future studies should directly address the potential causality of alterations in polyamine metabolism as well as associated pathways (such as arginine metabolism) for disease development and progression. It will be fundamental to test whether exogenously supplemented spermidine would counteract pathological polyamine biosynthesis activities (as indicated by increases in spermidine levels) through a direct or indirect metabolic feedback regulation. Mechanistically, such feedback effects may thus overcome maladaptive changes in the associated pathways (arginine and methionine or SAM metabolism) and define a protective "metabolic" action of dietary spermidine.

Conclusions and open questions

Dietary supplementation of spermidine prolongs life span and health span by protecting from a range of age-associated pathologies in several animal models. Although spermidine induces autophagy and autophagy inhibition curtails most of the spermidine effects, alternative mechanisms have also been implicated in the health-improving action of this polyamine. These potentially autophagy-independent mechanisms include direct antioxidant and metabolic effects on polyamine and associated metabolic pathways—in particular, increased arginine bioavailability and NO production (*44*, *65*). However, it has not been formally determined whether these routes act in a completely autophagy-independent manner or are interrelated with autophagy (in an additive or synergistic way), and it will be important to define actionable molecular targets that explain the beneficial effects of spermidine in diverse pathophysiological settings. In this sense, it will also be of interest to explore synergisms of CRMs that initially act through different mechanisms. For instance, the combination of spermidine and resveratrol shows synergistic effects on autophagy induction (*19*). Combinations with other CRMs, as well as with agents that act independently from protein acetylation such as rapamycin (table S1), remain to be tested. Mechanistic relations of spermidine to the autophagy inducer coffee may also exist (*129*). Coffee consumption inversely associates with total and cause-specific mortality in humans (*130*), and coffee induces autophagy and cytoplasmic protein deacetylation in mice (*129*). Although it appears that both effects are independent from caffeine content, the active principle(s) of coffee must be defined in molecular terms. It also remains to be determined whether combinations of distinct autophagy inducers can be used to obtain additive or synergistic health benefits or whether they rather would trigger unphysiological, potentially toxic levels of autophagy. Another important question is the tissue specificity of spermidine-induced health effects. Which are the prime targets of spermidine in aging and disease? Although cardiac and vascular tissues appear to be directly targeted, the kidney may similarly benefit from increased spermidine bioavailability because spermidine-rich diets ameliorate the functional decline of kidneys in animal models of aging and hypertension (*44*, *52*). The origin of spermidine-induced systemic effects on blood metabolites and proteins is another demanding problem. It remains unknown whether spermidine acts solely on leukocytes to suppress chronic low-grade inflammation and which organ is responsible for the increased bioavailability of arginine upon spermidine supplementation.

One of the strongest arguments for spermidine as a favorable CRM for future clinical trials is its low toxicity yet strong efficacy, even at moderate concentrations. Spermidine is an abundant natural polyamine contained in all organisms from bacteria to men and is naturally present in reasonable but varying amounts in human diets. It is therefore not surprising that life-long supplementation of spermidine does not seem to have any negative side effects in mice (*44*). Enhancing spermidine uptake could be achieved with a variety of strategies in clinical studies, namely by (i) supplementation of synthetic spermidine, (ii) changes in diet composition in favor of polyamine-rich food items, (iii) use of natural plant extracts rich in polyamines, or (iv) administration of prebiotics and probiotics that favor microbial polyamine synthesis in the gut. A polyamine-rich diet including daily servings of natto was well tolerated for a period of 2 months in human volunteers (*40*) and led to an effective increase in circulating spermine levels. A clinical trial on supplementation of spermidine-rich plant extracts to elderly is currently ongoing (https://clinicaltrials.gov/ct2/show/NCT02755246) (*131*). Nevertheless, possible contraindications of spermidine administration have to be carefully defined in clinical trials and may include patients with advanced cancer and renal failure. A low-polyamine diet may also be beneficial in some specific cases, as has been suggested in rats infected with *Trypanosoma brucei brucei* (*132*). Last, it may be interesting to develop spermidine analogs that act at lower doses or have a more stably pharmacokinetic profile for the prevention or treatment of specific diseases.

REFERENCES AND NOTES

1. R. de Cabo, D. Carmona-Gutierrez, M. Bernier, M. N. Hall, F. Madeo, The search for antiaging interventions: From elixirs to fasting regimens. *Cell* **157**, 1515–1526 (2014). doi: 10.1016/j.cell.2014.05.031; pmid: 24949965
2. C. López-Otín, L. Galluzzi, J. M. P. Freije, F. Madeo, G. Kroemer, Metabolic control of longevity. *Cell* **166**, 802–821 (2016). doi: 10.1016/j.cell.2016.07.031; pmid: 27518560
3. L. Fontana, L. Partridge, V. D. Longo, Extending healthy life span—From yeast to humans. *Science* **328**, 321–326 (2010). doi: 10.1126/science.1172539; pmid: 20395504
4. V. D. Longo, S. Panda, Fasting, Circadian rhythms, and time-restricted feeding in healthy lifespan. *Cell Metab.* **23**, 1048–1059 (2016). doi: 10.1016/j.cmet.2016.06.001; pmid: 27304506
5. R. M. Anson *et al.*, Intermittent fasting dissociates beneficial effects of dietary restriction on glucose metabolism and neuronal resistance to injury from calorie intake. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6216–6220 (2003). doi: 10.1073/pnas.1035720100; pmid: 12724520
6. R. R. Wing, S. Phelan, Long-term weight loss maintenance. *Am. J. Clin. Nutr.* **82** (Suppl), 222S–225S (2005). pmid: 16002825
7. F. Madeo, F. Pietrocola, T. Eisenberg, G. Kroemer, Caloric restriction mimetics: Towards a molecular definition. *Nat. Rev. Drug Discov.* **13**, 727–740 (2014). doi: 10.1038/nrd4391; pmid: 25212602
8. C. López-Otín, M. A. Blasco, L. Partridge, M. Serrano, G. Kroemer, The hallmarks of aging. *Cell* **153**, 1194–1217 (2013). doi: 10.1016/j.cell.2013.05.039; pmid: 23746838
9. C. Franceschi *et al.*, Inflamm-aging. An evolutionary perspective on immunosenescence. *Ann. N. Y. Acad. Sci.* **908**, 244–254 (2000). doi: 10.1111/j.1749-6632.2000.tb06651.x; pmid: 10911963
10. R. K. Junnila, E. O. List, D. E. Berryman, J. W. Murrey, J. J. Kopchick, The GH/IGF-1 axis in ageing and longevity. *Nat. Rev. Endocrinol.* **9**, 366–376 (2013). doi: 10.1038/nrendo.2013.67; pmid: 23591370
11. J. L. Barger *et al.*, Identification of tissue-specific transcriptional markers of caloric restriction in the mouse and their use to evaluate caloric restriction mimetics. *Aging Cell* **16**, 750–760 (2017). doi: 10.1111/acel.12608; pmid: 28556428
12. D. C. Rubinsztein, G. Mariño, G. Kroemer, Autophagy and aging. *Cell* **146**, 682–695 (2011). doi: 10.1016/j.cell.2011.07.030; pmid: 21884931
13. R. A. Frake, T. Ricketts, F. M. Menzies, D. C. Rubinsztein, Autophagy and neurodegeneration. *J. Clin. Invest.* **125**, 65–74 (2015). doi: 10.1172/JCI73944; pmid: 25654552
14. A. M. K. Choi, S. W. Ryter, B. Levine, Autophagy in human health and disease. *N. Engl. J. Med.* **368**, 651–662 (2013). doi: 10.1056/NEJMra1205406; pmid: 23406030
15. C. Zhang, A. M. Cuervo, Restoration of chaperone-mediated autophagy in aging liver improves cellular maintenance and hepatic function. *Nat. Med.* **14**, 959–965 (2008). doi: 10.1038/nm.1851; pmid: 18690243
16. T. Eisenberg *et al.*, Nucleocytosolic depletion of the energy metabolite acetyl-coenzyme a stimulates autophagy and prolongs lifespan. *Cell Metab.* **19**, 431–444 (2014). doi: 10.1016/j.cmet.2014.02.010; pmid: 24606900

17. G. Mariño *et al*., Regulation of autophagy by cytosolic acetyl-coenzyme A. *Mol. Cell* **53**, 710–725 (2014). doi: 10.1016/j.molcel.2014.01.016; pmid: 24560926
18. F. Pietrocola *et al*., Spermidine induces autophagy by inhibiting the acetyltransferase EP300. *Cell Death Differ.* **22**, 509–516 (2015). doi: 10.1038/cdd.2014.215; pmid: 25526088
19. E. Morselli *et al*., Spermidine and resveratrol induce autophagy by distinct pathways converging on the acetylproteome. *J. Cell Biol.* **192**, 615–629 (2011). doi: 10.1083/jcb.201008167; pmid: 21339330
20. A. E. Pegg, Functions of polyamines in mammals. *J. Biol. Chem.* **291**, 14904–14912 (2016). doi: 10.1074/jbc.R116.731661; pmid: 27268251
21. S. Bardócz *et al*., The importance of dietary polyamines in cell regeneration and growth. *Br. J. Nutr.* **73**, 819–828 (1995). doi: 10.1079/BJN19950087; pmid: 7632663
22. K. Igarashi, K. Kashiwagi, Modulation of cellular function by polyamines. *Int. J. Biochem. Cell Biol.* **42**, 39–51 (2010). doi: 10.1016/j.biocel.2009.07.009; pmid: 19643201
23. G. P. Guerra, M. A. Rubin, C. F. Mello, Modulation of learning and memory by natural polyamines. *Pharmacol. Res.* **112**, 99–118 (2016). doi: 10.1016/j.phrs.2016.03.023; pmid: 27015893
24. T. Eisenberg *et al*., Induction of autophagy by spermidine promotes longevity. *Nat. Cell Biol.* **11**, 1305–1314 (2009). doi: 10.1038/ncb1975; pmid: 19801973
25. S. Pucciarelli *et al*., Spermidine and spermine are enriched in whole blood of nona/centenarians. *Rejuvenation Res.* **15**, 590–595 (2012). doi: 10.1089/rej.2012.1349; pmid: 22950434
26. V. K. Gupta *et al*., Restoring polyamines protects from age-induced memory impairment in an autophagy-dependent manner. *Nat. Neurosci.* **16**, 1453–1460 (2013). doi: 10.1038/nn.3512; pmid: 23995066
27. G. Scalabrino, M. E. Ferioli, Polyamines in mammalian ageing: An oncological problem, too? A review. *Mech. Ageing Dev.* **26**, 149–164 (1984). doi: 10.1016/0047-6374(84)90090-3; pmid: 6384679
28. K. Nishimura, R. Shiina, K. Kashiwagi, K. Igarashi, Decrease in polyamines with aging and their ingestion from food and drink. *J. Biochem.* **139**, 81–90 (2006). doi: 10.1093/jb/mvj003; pmid: 16428322
29. A. E. Pegg, Mammalian polyamine metabolism and function. *IUBMB Life* **61**, 880–894 (2009). doi: 10.1002/iub.230; pmid: 19603518
30. K. Kee *et al*., Activated polyamine catabolism depletes acetyl-CoA pools and suppresses prostate tumor growth in TRAMP mice. *J. Biol. Chem.* **279**, 40076–40083 (2004). doi: 10.1074/jbc.M406002200; pmid: 15252047
31. J. Jell *et al*., Genetically altered expression of spermidine/spermine N1-acetyltransferase affects fat metabolism in mice via acetyl-CoA. *J. Biol. Chem.* **282**, 8404–8413 (2007). doi: 10.1074/jbc.M610265200; pmid: 17189273
32. I. Oehme *et al*., Histone deacetylase 10 promotes autophagy-mediated cell survival. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E2592–E2601 (2013). doi: 10.1073/pnas.1300113110; pmid: 23801752
33. Y. Hai, S. A. Shinsky, N. J. Porter, D. W. Christianson, Histone deacetylase 10 structure and molecular function as a polyamine deacetylase. *Nat. Commun.* **8**, 15368 (2017). doi: 10.1038/ncomms15368; pmid: 28516954
34. A. B. Knott, E. Bossy-Wetzel, Impact of nitric oxide on metabolism in health and age-related disease. *Diabetes Obes. Metab.* **12** (suppl. 2), 126–133 (2010). doi: 10.1111/j.1463-1326.2010.01267.x; pmid: 21029309
35. S. L. Nowotarski, P. M. Woster, R. A. Casero Jr., Polyamines and cancer: Implications for chemotherapy and chemoprevention. *Expert Rev. Mol. Med.* **15**, e3 (2013). doi: 10.1017/erm.2013.3; pmid: 23432971
36. M. Hiasa *et al*., Identification of a mammalian vesicular polyamine transporter. *Sci. Rep.* **4**, 6836 (2014). doi: 10.1038/srep06836; pmid: 25355561
37. T. Takeuchi *et al*., Vesicular polyamine transporter mediates vesicular storage and release of polyamine from mast cells. *J. Biol. Chem.* **292**, 3909–3918 (2017). doi: 10.1074/jbc.M116.756197; pmid: 28082679
38. T. Masuko *et al*., Polyamine transport, accumulation, and release in brain. *J. Neurochem.* **84**, 610–617 (2003). doi: 10.1046/j.1471-4159.2003.01558.x; pmid: 12558981
39. V. Milovic, Polyamines in the gut lumen: Bioavailability and biodistribution. *Eur. J. Gastroenterol. Hepatol.* **13**, 1021–1025 (2001). doi: 10.1097/00042737-200109000-00004; pmid: 11564949
40. K. Soda *et al*., Long-term oral polyamine intake increases blood polyamine concentrations. *J. Nutr. Sci. Vitaminol. (Tokyo)* **55**, 361–366 (2009). doi: 10.3177/jnsv.55.361; pmid: 19763038
41. M. Atiya Ali, E. Poortvliet, R. Strömberg, A. Yngve, Polyamines in foods: Development of a food database. *Food Nutr. Res.* **55**, (2011). doi: 10.3402/fnr.v55i0.5572; pmid: 21249159
42. C. Zoumas-Morse *et al*., Development of a polyamine database for assessing dietary intake. *J. Am. Diet. Assoc.* **107**, 1024–1027 (2007). doi: 10.1016/j.jada.2007.03.012; pmid: 17524725
43. M. A. Ali, E. Poortvliet, R. Strömberg, A. Yngve, Polyamines: Total daily intake in adolescents compared to the intake estimated from the Swedish Nutrition Recommendations Objectified (SNO). *Food Nutr. Res.* **55**, 5455 (2011). doi: 10.3402/fnr.v55i0.5455; pmid: 21249160
44. T. Eisenberg *et al*., Cardioprotection and lifespan extension by the natural polyamine spermidine. *Nat. Med.* **22**, 1428–1438 (2016). doi: 10.1038/nm.4222; pmid: 27841876
45. Z. Zwighaft *et al*., Circadian clock control by polyamine levels through a mechanism that declines with age. *Cell Metab.* **22**, 874–885 (2015). doi: 10.1016/j.cmet.2015.09.011; pmid: 26456331
46. R. Kibe *et al*., Upregulation of colonic luminal polyamines produced by intestinal microbiota delays senescence in mice. *Sci. Rep.* **4**, 4548 (2014). doi: 10.1038/srep04548; pmid: 24686447
47. M. Matsumoto, S. Kurihara, R. Kibe, H. Ashida, Y. Benno, Longevity in mice is promoted by probiotic-induced suppression of colonic senescence dependent on upregulation of gut bacterial polyamine production. *PLOS ONE* **6**, e23652 (2011). doi: 10.1371/journal.pone.0023652; pmid: 21858192
48. M. Matsumoto *et al*., Impact of intestinal microbiota on intestinal luminal metabolome. *Sci. Rep.* **2**, 233 (2012). doi: 10.1038/srep00233; pmid: 22724057
49. M. Matsumoto, Y. Benno, The relationship between microbiota and polyamine concentration in the human intestine: A pilot study. *Microbiol. Immunol.* **51**, 25–35 (2007). doi: 10.1111/j.1348-0421.2007.tb03887.x; pmid: 17237596
50. K. Igarashi, S. Ueda, K. Yoshida, K. Kashiwagi, Polyamines in renal failure. *Amino Acids* **31**, 477–483 (2006). doi: 10.1007/s00726-006-0264-7; pmid: 16554974
51. F. Yue *et al*., Spermidine prolongs lifespan and prevents liver fibrosis and hepatocellular carcinoma by activating MAP1S-mediated autophagy. *Cancer Res.* **77**, 2938–2951 (2017). doi: 10.1158/0008-5472.CAN-16-3462; pmid: 28386016
52. K. Soda, Y. Dobashi, Y. Kano, S. Tsujinaka, F. Konishi, Polyamine-rich food decreases age-associated pathology and mortality in aged mice. *Exp. Gerontol.* **44**, 727–732 (2009). doi: 10.1016/j.exger.2009.08.013; pmid: 19735716
53. T. Yamamoto *et al*., The natural polyamines spermidine and spermine prevent bone loss through preferential disruption of osteoclastic activation in ovariectomized mice. *Br. J. Pharmacol.* **166**, 1084–1096 (2012). doi: 10.1111/j.1476-5381.2012.01856.x; pmid: 22250848
54. P. N. T. Binh, K. Soda, C. Maruyama, M. Kawakami, Relationship between food polyamines and gross domestic product in association with longevity in Asian countries. *Health (N. Y.)* **02**, 1390 (2010).
55. E. W. Gerner, F. L. Meyskens Jr., Polyamines and cancer: Old molecules, new understanding. *Nat. Rev. Cancer* **4**, 781–792 (2004). doi: 10.1038/nrc1454; pmid: 15510159
56. M. Pless *et al*., Clinical efficacy, tolerability, and safety of SAM486A, a novel polyamine biosynthesis inhibitor, in patients with relapsed or refractory non-Hodgkin's lymphoma: Results from a phase II multicenter study. *Clin. Cancer Res.* **10**, 1299–1305 (2004). doi: 10.1158/1078-0432.CCR-0977-03; pmid: 14977828
57. H. Miao *et al*., Macrophage ABHD5 promotes colorectal cancer growth by suppressing spermidine production by SRM. *Nat. Commun.* **7**, 11716 (2016). doi: 10.1038/ncomms11716; pmid: 27189574
58. K. Soda, Y. Kano, F. Chiba, K. Koizumi, Y. Miyaki, Increased polyamine intake inhibits age-associated alteration in global DNA methylation and 1,2-dimethylhydrazine-induced tumorigenesis. *PLOS ONE* **8**, e64357 (2013). doi: 10.1371/journal.pone.0064357; pmid: 23696883
59. A. J. Vargas *et al*., Dietary polyamine intake and risk of colorectal adenomatous polyps. *Am. J. Clin. Nutr.* **96**, 133–141 (2012). doi: 10.3945/ajcn.111.030353; pmid: 22648715
60. A. J. Vargas *et al*., Dietary polyamine intake and colorectal cancer risk in postmenopausal women. *Am. J. Clin. Nutr.* **102**, 411–419 (2015). doi: 10.3945/ajcn.114.103895; pmid: 26135350
61. F. Pietrocola *et al*., Caloric restriction mimetics enhance anticancer immunosurveillance. *Cancer Cell* **30**, 147–160 (2016). doi: 10.1016/j.ccell.2016.05.016; pmid: 27411589
62. C. Lee *et al*., Reduced levels of IGF-I mediate differential protection of normal and cancer cells in response to fasting and improve chemotherapeutic index. *Cancer Res.* **70**, 1564–1572 (2010). doi: 10.1158/0008-5472.CAN-09-3228; pmid: 20145127
63. S. Di Biase *et al*., Fasting-mimicking diet reduces HO-1 to promote T cell-mediated tumor cytotoxicity. *Cancer Cell* **30**, 136–146 (2016). doi: 10.1016/j.ccell.2016.06.005; pmid: 27411588
64. K. Soda, Y. Kano, F. Chiba, Food polyamine and cardiovascular disease—An epidemiological study. *Glob. J. Health Sci.* **4**, 170–178 (2012). doi: 10.5539/gjhs.v4n6p170; pmid: 23121753
65. T. J. LaRocca, R. A. Gioscia-Ryan, C. M. Hearon Jr., D. R. Seals, The autophagy enhancer spermidine reverses arterial aging. *Mech. Ageing Dev.* **134**, 314–320 (2013). doi: 10.1016/j.mad.2013.04.004; pmid: 23612189
66. C. F. Michiels, A. Kurdi, J.-P. Timmermans, G. R. Y. De Meyer, W. Martinet, Spermidine reduces lipid accumulation and necrotic core formation in atherosclerotic plaques via induction of autophagy. *Atherosclerosis* **251**, 319–327 (2016). doi: 10.1016/j.atherosclerosis.2016.07.899; pmid: 27450786
67. T. Hayashi *et al*., Decreased ornithine decarboxylase activity in the kidneys of Dahl salt-sensitive rats. *Hypertens. Res.* **25**, 787–795 (2002). doi: 10.1291/hypres.25.787; pmid: 12452334
68. T. Chen *et al*., Rapamycin and other longevity-promoting compounds enhance the generation of mouse induced pluripotent stem cells. *Aging Cell* **10**, 908–911 (2011). doi: 10.1111/j.1474-9726.2011.00722.x; pmid: 21615676
69. Y. Ramot *et al*., Spermidine promotes human hair growth and is a novel modulator of human epithelial stem cell functions. *PLOS ONE* **6**, e22564 (2011). doi: 10.1371/journal.pone.0022564; pmid: 21818338
70. L. García-Prat *et al*., Autophagy maintains stemness by preventing senescence. *Nature* **529**, 37–42 (2016). doi: 10.1038/nature16187; pmid: 26738589
71. M. Chrisam *et al*., Reactivation of autophagy by spermidine ameliorates the myopathic defects of collagen VI-null mice. *Autophagy* **11**, 2142–2152 (2015). doi: 10.1080/15548627.2015.1108508; pmid: 26565691
72. J. Fan *et al*., Spermidine coupled with exercise rescues skeletal muscle atrophy from D-gal-induced aging rats through enhanced autophagy and reduced apoptosis via AMPK-FOXO3a signal pathway. *Oncotarget* **8**, 17475–17490 (2017). pmid: 28407698
73. S. K. Sadasivan *et al*., Exogenous administration of spermine improves glucose utilization and decreases bodyweight in mice. *Eur. J. Pharmacol.* **729**, 94–99 (2014). doi: 10.1016/j.ejphar.2014.01.073; pmid: 24530553
74. D. Kraus *et al*., Nicotinamide *N*-methyltransferase knockdown protects against diet-induced obesity. *Nature* **508**, 258–262 (2014). doi: 10.1038/nature13198; pmid: 24717514
75. N. Bonhoure *et al*., Loss of the RNA polymerase III repressor MAF1 confers obesity resistance. *Genes Dev.* **29**, 934–947 (2015). doi: 10.1101/gad.258350.115; pmid: 25934505
76. Á. F. Fernández *et al*., Autophagy couteracts weight gain, lipotoxicity and pancreatic β-cell death upon hypercaloric pro-diabetic regimens. *Cell Death Dis.* **8**, e2970 (2017). doi: 10.1038/cddis.2017.373; pmid: 28771229
77. M. T. Lin, M. F. Beal, Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. *Nature* **443**, 787–795 (2006). doi: 10.1038/nature05292; pmid: 17051205
78. N. Minois, P. Rockenfeller, T. K. Smith, D. Carmona-Gutierrez, Spermidine feeding decreases age-related locomotor activity loss and induces changes in lipid composition. *PLOS ONE* **9**, e102435 (2014). doi: 10.1371/journal.pone.0102435; pmid: 25010732
79. V. K. Gupta *et al*., Spermidine suppresses age-associated memory impairment by preventing adverse increase of presynaptic active zone size and release. *PLOS Biol.* **14**, e1002563 (2016). doi: 10.1371/journal.pbio.1002563; pmid: 27684064

80. X. Guo *et al*., Spermidine alleviates severity of murine experimental autoimmune encephalomyelitis. *Invest. Opthalmol. Vis. Sci.* **52**, 2696–2703 (2011). doi: 10.1167/iovs.10-6015; pmid: 21228387
81. Q. Yang *et al*., Spermidine alleviates experimental autoimmune encephalomyelitis through inducing inhibitory macrophages. *Cell Death Differ.* **23**, 1850–1861 (2016). doi: 10.1038/cdd.2016.71; pmid: 27447115
82. T. Noro *et al*., Spermidine promotes retinal ganglion cell survival and optic nerve regeneration in adult mice following optic nerve injury. *Cell Death Dis.* **6**, e1720 (2015). doi: 10.1038/cddis.2015.93; pmid: 25880087
83. T. Noro *et al*., Spermidine ameliorates neurodegeneration in a mouse model of normal tension glaucoma. *Invest. Ophthalmol. Vis. Sci.* **56**, 5012–5019 (2015). doi: 10.1167/iovs.15-17142; pmid: 26230766
84. K. Deng *et al*., Increased synthesis of spermidine as a result of upregulation of arginase I promotes axonal regeneration in culture and in vivo. *J. Neurosci.* **29**, 9545–9552 (2009). doi: 10.1523/JNEUROSCI.1175-09.2009; pmid: 19641117
85. S. Büttner *et al*., Spermidine protects against α-synuclein neurotoxicity. *Cell Cycle* **13**, 3903–3908 (2014). doi: 10.4161/15384101.2014.973309; pmid: 25483063
86. I.-F. Wang *et al*., Autophagy activators rescue and alleviate pathogenesis of a mouse model with proteinopathies of the TAR DNA-binding protein 43. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 15024–15029 (2012). doi: 10.1073/pnas.1206362109; pmid: 22932872
87. Y. Zhang *et al*., Spermidine preconditioning ameliorates laurate-induced brain injury by maintaining mitochondrial stability. *Neurol. Res.* **39**, 248–258 (2017). doi: 10.1080/01616412.2017.1283830; pmid: 28112032
88. N. A. Velloso *et al*., Spermine improves recognition memory deficit in a rodent model of Huntington's disease. *Neurobiol. Learn. Mem.* **92**, 574–580 (2009). doi: 10.1016/j.nlm.2009.07.006; pmid: 19632348
89. Y. Qi, Q. Qiu, X. Gu, Y. Tian, Y. Zhang, ATM mediates spermidine-induced mitophagy via PINK1 and Parkin regulation in human fibroblasts. *Sci. Rep.* **6**, 24700 (2016). doi: 10.1038/srep24700; pmid: 27089984
90. G. Mariño, F. Pietrocola, F. Madeo, G. Kroemer, Caloric restriction mimetics: Natural/physiological pharmacological autophagy inducers. *Autophagy* **10**, 1879–1882 (2014). doi: 10.4161/auto.36413; pmid: 25484097
91. N. Minois *et al*., Spermidine promotes stress resistance in Drosophila melanogaster through autophagy-dependent and -independent pathways. *Cell Death Dis.* **3**, e401 (2012). doi: 10.1038/cddis.2012.139; pmid: 23059820
92. I. H. Lee, T. Finkel, Regulation of autophagy by the p300 acetyltransferase. *J. Biol. Chem.* **284**, 6322–6328 (2009). doi: 10.1074/jbc.M807135200; pmid: 19124466
93. S. Sebti *et al*., BAT3 modulates p300-dependent acetylation of p53 and autophagy-related protein 7 (ATG7) during autophagy. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 4115–4120 (2014). doi: 10.1073/pnas.1313618111; pmid: 24591579
94. R. Mackeh *et al*., Reactive oxygen species, AMP-activated protein kinase, and the transcription cofactor p300 regulate α-tubulin acetyltransferase-1 (αTAT-1/MEC-17)-dependent microtubule hyperacetylation during cell stress. *J. Biol. Chem.* **289**, 11816–11828 (2014). doi: 10.1074/jbc.M113.507400; pmid: 24619423
95. C. Mammucari *et al*., FoxO3 controls autophagy in skeletal muscle in vivo. *Cell Metab.* **6**, 458–471 (2007). doi: 10.1016/j.cmet.2007.11.001; pmid: 18054315
96. F. Pietrocola, L. Galluzzi, J. M. Bravo-San Pedro, F. Madeo, G. Kroemer, Acetyl coenzyme A: A central metabolite and second messenger. *Cell Metab.* **21**, 805–821 (2015). doi: 10.1016/j.cmet.2015.05.014; pmid: 26039447
97. H. S. Ghosh, M. McBurney, P. D. Robbins, SIRT1 negatively regulates the mammalian target of rapamycin. *PLOS ONE* **5**, e9199 (2010). doi: 10.1371/journal.pone.0009199; pmid: 20169165
98. A. Eiyama, K. Okamoto, PINK1/Parkin-mediated mitophagy in mammalian cells. *Curr. Opin. Cell Biol.* **33**, 95–101 (2015). doi: 10.1016/j.ceb.2015.01.002; pmid: 25697963
99. M. V. Bennetzen *et al*., Phosphoproteomic analysis of cells treated with longevity-related autophagy inducers. *Cell Cycle Georget. Tex.* **11**, 1827–1840 (2012).
100. E. Vacchelli *et al*., Chemotherapy-induced antitumor immunity requires formyl peptide receptor 1. *Science* **350**, 972–978 (2015). doi: 10.1126/science.aad0779; pmid: 26516201
101. L. Galluzzi, A. Buqué, O. Kepp, L. Zitvogel, G. Kroemer, Immunogenic cell death in cancer and infectious disease. *Nat. Rev. Immunol.* **17**, 97–111 (2017). doi: 10.1038/nri.2016.107; pmid: 27748397
102. L. Galluzzi, A. Buqué, O. Kepp, L. Zitvogel, G. Kroemer, Immunological effects of conventional chemotherapy and targeted anticancer agents. *Cancer Cell* **28**, 690–714 (2015). doi: 10.1016/j.ccell.2015.10.012; pmid: 26678337
103. Y. Ma *et al*., Anticancer chemotherapy-induced intratumoral recruitment and differentiation of antigen-presenting cells. *Immunity* **38**, 729–741 (2013). doi: 10.1016/j.immuni.2013.03.003; pmid: 23562161
104. M. Michaud *et al*., Autophagy-dependent anticancer immune responses induced by chemotherapeutic agents in mice. *Science* **334**, 1573–1577 (2011). doi: 10.1126/science.1208347; pmid: 22174255
105. S. Rao *et al*., A dual role for autophagy in a murine model of lung cancer. *Nat. Commun.* **5**, 3056 (2014). doi: 10.1038/ncomms4056; pmid: 24445999
106. C. Lee *et al*., Fasting cycles retard growth of tumors and sensitize a range of cancer cell types to chemotherapy. *Sci. Transl. Med.* **4**, 124ra27 (2012). doi: 10.1126/scitranslmed.3003293; pmid: 22323820
107. K. Araki *et al*., mTOR regulates memory CD8 T-cell differentiation. *Nature* **460**, 108–112 (2009). doi: 10.1038/nature08155; pmid: 19543266
108. E. L. Pearce *et al*., Enhancing CD8 T-cell memory by modulating fatty acid metabolism. *Nature* **460**, 103–107 (2009). doi: 10.1038/nature08097; pmid: 19494812
109. K. Phadwal *et al*., A novel method for autophagy detection in primary cells: Impaired levels of macroautophagy in immunosenescent T cells. *Autophagy* **8**, 677–689 (2012). doi: 10.4161/auto.18935; pmid: 22302009
110. D. J. Puleston *et al*., Autophagy is a critical regulator of memory $CD8^+$ T cell formation. *eLife* **3**, e03706 (2014). doi: 10.7554/eLife.03706; pmid: 25385531
111. D. J. Puleston, A. K. Simon, New roles for autophagy and spermidine in T cells. *Microb. Cell* **2**, 91–93 (2015). doi: 10.15698/mic2015.03.195; pmid: 28357282
112. Z. Zhong, E. Sanchez-Lopez, M. Karin, Autophagy, NLRP3 inflammasome and auto-inflammatory/immune diseases. *Clin. Exp. Rheumatol.* **34** (Suppl 98), 12–16 (2016). pmid: 27586797
113. S. Zhu *et al*., Spermine protects mice against lethal sepsis partly by attenuating surrogate inflammatory markers. *Mol. Med.* **15**, 275–282 (2009). pmid: 19593412
114. K. Soda *et al*., Spermine, a natural polyamine, suppresses LFA-1 expression on human lymphocyte. *J. Immunol.* **175**, 237–245 (2005). doi: 10.4049/jimmunol.175.1.237; pmid: 15972654
115. Y. Kano, K. Soda, F. Konishi, Suppression of LFA-1 expression by spermine is associated with enhanced methylation of ITGAL, the LFA-1 promoter area. *PLOS ONE* **8**, e56056 (2013). doi: 10.1371/journal.pone.0056056; pmid: 23418509
116. W. W. Shin, W. F. Fong, S. F. Pang, P. C. Wong, Limited blood-brain barrier transport of polyamines. *J. Neurochem.* **44**, 1056–1059 (1985). doi: 10.1111/j.1471-4159.1985.tb08724.x; pmid: 3973604
117. A. S. Diler *et al*., Passage of spermidine across the blood-brain barrier in short recirculation periods following global cerebral ischemia: Effects of mild hyperthermia. *Neurosci. Res.* **43**, 335–342 (2002). doi: 10.1016/S0168-0102(02)00059-7; pmid: 12135777
118. E. Sahin *et al*., Macrophage PTEN regulates expression and secretion of arginase I modulating innate and adaptive immune responses. *J. Immunol.* **193**, 1717–1727 (2014). doi: 10.4049/jimmunol.1302167; pmid: 25015834
119. S. Jo *et al*., GABA from reactive astrocytes impairs memory in mouse models of Alzheimer's disease. *Nat. Med.* **20**, 886–896 (2014). doi: 10.1038/nm.3639; pmid: 24973918
120. J. Saura *et al*., Increased monoamine oxidase B activity in plaque-associated astrocytes of Alzheimer brains revealed by quantitative enzyme radioautography. *Neuroscience* **62**, 15–30 (1994). doi: 10.1016/0306-4522(94)90311-5; pmid: 7816197
121. M. H. Park, K. Igarashi, Polyamines and their metabolites as diagnostic markers of human diseases. *Biomol. Ther. (Seoul)* **21**, 1–9 (2013). doi: 10.4062/biomolther.2012.097; pmid: 24009852
122. M.-L. Cheng *et al*., Metabolic disturbances identified in plasma are associated with outcomes in patients with heart failure: Diagnostic and prognostic value of metabolomics. *J. Am. Coll. Cardiol.* **65**, 1509–1520 (2015). doi: 10.1016/j.jacc.2015.02.018; pmid: 25881932
123. B. E. Sansbury *et al*., Metabolomic analysis of pressure-overloaded and infarcted mouse hearts. *Circ Heart Fail* **7**, 634–642 (2014). doi: 10.1161/CIRCHEARTFAILURE.114.001151; pmid: 24762972
124. X. Pan *et al*., Alzheimer's disease-like pathology has transient effects on the brain and blood metabolome. *Neurobiol. Aging* **38**, 151–163 (2016). doi: 10.1016/j.neurobiolaging.2015.11.014; pmid: 26827653
125. S. F. Graham *et al*., Untargeted metabolomic analysis of human plasma indicates differentially affected polyamine and L-arginine metabolism in mild cognitive impairment subjects converting to Alzheimer's disease. *PLOS ONE* **10**, e0119452 (2015). doi: 10.1371/journal.pone.0119452; pmid: 25803028
126. C. Magnes *et al*., Polyamines in biological samples: Rapid and robust quantification by solid-phase extraction online-coupled to liquid chromatography-tandem mass spectrometry. *J. Chromatogr. A* **1331**, 44–51 (2014). doi: 10.1016/j.chroma.2013.12.061; pmid: 24485539
127. M. Ost *et al*., Muscle mitohormesis promotes cellular survival via serine/glycine pathway flux. *FASEB J.* **29**, 1314–1328 (2015). doi: 10.1096/fj.14-261503; pmid: 25491309
128. P. Liu, N. Gupta, Y. Jing, H. Zhang, Age-related changes in polyamines in memory-associated brain structures in rats. *Neuroscience* **155**, 789–796 (2008). doi: 10.1016/j.neuroscience.2008.06.033; pmid: 18621105
129. F. Pietrocola *et al*., Coffee induces autophagy in vivo. *Cell Cycle* **13**, 1987–1994 (2014). doi: 10.4161/cc.28929; pmid: 24769862
130. N. D. Freedman, Y. Park, C. C. Abnet, A. R. Hollenbeck, R. Sinha, Association of coffee drinking with total and cause-specific mortality. *N. Engl. J. Med.* **366**, 1891–1904 (2012). doi: 10.1056/NEJMoa1112010; pmid: 22591295
131. C. Schwarz *et al*., Safety and tolerability of spermidine supplementation in mice and older adults with subjective cognitive decline. *Aging* (2018). doi: 10.18632/aging.101354; pmid: 29315079
132. K. Nishimura *et al*., Effect of polyamine-deficient chow on *Trypanosoma brucei brucei* infection in rats. *J. Parasitol.* **95**, 781–786 (2009). doi: 10.1645/GE-1883.1; pmid: 20049984

ACKNOWLEDGMENTS

F.M., T.E., and G.K. are inventors on a patent application (WO2010081204 A3) submitted by Katholieke Universiteit Leuven that covers the use of spermidine for life-span extension. F.P. and G.K. are inventors on a patent application (WO2015049365 A3) submitted by INSERM (Institut National de la Santé et de la Recherche Médicale), Assistance Publique-Hôpitaux De Paris (APHP), Université Paris Descartes, Université Pierre et Marie Curie (Paris 6), Université Paris Diderot–Paris 7, Université Paris–Sud, Institut Gustave Roussy, that covers the medical use of CRMs. We thank D. Carmona-Gutierrez and P. Royer for critical reading and support. F.M. is grateful to the Austrian Science Fund FWF (Austria) for grants P23490-B20, P29262, P24381, P29203 P27893, I1000, and "SFB Lipotox" (F3012), as well as to Bundesministerium für Wissenschaft, Forschung und Wirtschaft and the Karl-Franzens University for grant "Unkonventionelle Forschung" and grant DKplus Metabolic and Cardiovascular Diseases (W1226). We acknowledge support from NAWI Graz and the BioTechMed-Graz flagship project "EPIAge." G.K. is supported by the Ligue contre le Cancer Comité de Charente-Maritime (équipe labelisée); Agence National de la Recherche (ANR)–Projets blancs; ANR under the frame of E-Rare-2, the ERA-Net for Research on Rare Diseases; Association pour la recherche sur le cancer (ARC); Cancéropôle Ile-de-France; Institut National du Cancer (INCa); Institut Universitaire de France; Fondation pour la Recherche Médicale (FRM); the European Commission (ArtForce); the European Research Council (ERC); the LeDucq Foundation; the LabEx Immuno-Oncology; the Recherche Hospitalo-Universitaire Torino Lumière, the Site de Recherche Intégrée sur le Cancer (SIRIC) Stratified Oncology Cell DNA Repair and Tumor Immune Elimination (SOCRATE); the SIRIC Cancer Research and Personalized Medicine (CARPEM); and the Paris Alliance of Cancer Research Institutes (PACRI). T.E. is supported by the FFG COIN-project 855987. F.M. and T.E. have equity interests in TLL (The Longevity Labs), a company founded in 2016 that will develop natural food extracts.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6374/eaan2788/suppl/DC1
Tables S1 and S2
References (*133*–*275*)

10.1126/science.aan2788

RESEARCH ARTICLE

ATMOSPHERIC PHYSICS

Substantial convection and precipitation enhancements by ultrafine aerosol particles

Jiwen Fan,[1]* Daniel Rosenfeld,[2] Yuwei Zhang,[1,3] Scott E. Giangrande,[4] Zhanqing Li,[3,5] Luiz A. T. Machado,[6] Scot T. Martin,[7] Yan Yang,[1,8] Jian Wang,[4] Paulo Artaxo,[9] Henrique M. J. Barbosa,[9,10] Ramon C. Braga,[6] Jennifer M. Comstock,[1] Zhe Feng,[1] Wenhua Gao,[1,11] Helber B. Gomes,[12] Fan Mei,[1] Christopher Pöhlker,[13] Mira L. Pöhlker,[13] Ulrich Pöschl,[13,14] Rodrigo A. F. de Souza[15]

Aerosol-cloud interactions remain the largest uncertainty in climate projections. Ultrafine aerosol particles smaller than 50 nanometers ($UAP_{<50}$) can be abundant in the troposphere but are conventionally considered too small to affect cloud formation. Observational evidence and numerical simulations of deep convective clouds (DCCs) over the Amazon show that DCCs forming in a low-aerosol environment can develop very large vapor supersaturation because fast droplet coalescence reduces integrated droplet surface area and subsequent condensation. $UAP_{<50}$ from pollution plumes that are ingested into such clouds can be activated to form additional cloud droplets on which excess supersaturation condenses and forms additional cloud water and latent heating, thus intensifying convective strength. This mechanism suggests a strong anthropogenic invigoration of DCCs in previously pristine regions of the world.

Deep convective cloud (DCC) systems in the tropics produce copious precipitation and drive the global-scale circulation (*1*). Precipitation, latent heating, and cloud radiative forcing associated with DCCs are strongly modulated by cloud microphysical processes (*2*). These processes in tropical DCCs are initiated from droplet nucleation (which is determined by vapor supersaturation in updrafts and aerosol properties such as composition and size distribution. Aerosol impacts on cloud processes via this pathway are known as aerosol indirect effects, referred to as aerosol-cloud interaction in the most recent Intergovernmental Panel on Climate Change report (*3*). Aerosol impacts are a key uncertainty in understanding the current and future climate (*3*) as well as extreme weather (*4*). DCCs have complicated dynamics and microphysics; therefore, aerosol impacts on them are extremely complex and hard to disentangle. Previous studies have shown that aerosols could invigorate or suppress DCC intensity through aerosol indirect effects, contingent on dynamical and thermodynamical conditions [e.g., (*4*–*12*)]. In the case of warm-cloud bases (>15°C), increasing aerosol concentrations can suppress warm rain because of a reduction in droplet size, which allows more cloud water to be lifted to a higher altitude; the freezing of this larger amount of cloud water releases additional latent heat, thereby invigorating convective updrafts [referred to as "cold-phase invigoration" (*13*)]. The importance of the process of enhanced condensation by aerosols to deep-cloud development has been shown in tropical maritime clouds (*14*–*16*). Enhancement in DCC intensity favors enhanced storm electrification, larger precipitation rates, and taller clouds with larger anvils.

Over the Amazon, unperturbed background aerosol concentrations found within the region of pristine rainforests are low, with values reported in the low hundreds per cubic centimeter (*17*–*19*), similar to conditions thought to have existed in preindustrial times (*18*). Responses of cloud properties to aerosols are nonlinear and most sensitive to the addition of particles when the background concentration is very low [i.e., the "aerosol-limited regime" (*17*, *20*, *21*)]. Therefore, the aerosol effect on convective clouds over the Amazon region is expected to be large, and previous studies have demonstrated evident smoke aerosol influence on convective clouds during dry seasons in the Amazon (June to September) (*22*–*24*). In addition to the low total aerosol concentration, ultrafine aerosol particles with diameters less than 50 nm ($UAP_{<50}$) are typically nearly absent over the Amazon rainforest, as new particle formation has rarely been observed in the boundary layer there (*25*, *26*). $UAP_{<50}$ are often transported downward from the upper troposphere and subsequently grow to the particles larger than 50 nm that are traditionally considered as cloud condensation nuclei ($CCN_{>50}$) (*27*). However, the Manaus metropolis, a city of about 2 million people, produces a pollution plume that generally follows the northeasterly trade winds and is an important source of elevated $UAP_{<50}$ concentrations (*28*).

The Observations and Modeling of the Green Ocean Amazon (GoAmazon 2014/5) experiment was carried out near the Manaus metropolis to gain a better understanding of the impacts of pollution emissions from Manaus on the hydrological cycle and climate in the tropical rainforest (*29*, *30*). We took advantage of the unique observational data from GoAmazon 2014/5 (e.g., the direct cloud updraft velocity measurements for DCCs occurring in a similar convective environment but different aerosol environment) to investigate the observed aerosol effects, as well as to perform and analyze high-resolution simulations of a sample case, using the detailed spectral-bin microphysics scheme to scrutinize the mechanism. We found that the $UAP_{<50}$ introduced by the Manaus pollution plume enhanced convective intensity and precipitation rates to a degree not previously observed or simulated. The detailed simulations show that the drastic enhancement in convective intensity is primarily attributable to the enhanced condensational heating, with the latent heat released from enhanced ice-related processes at upper levels playing a secondary role. This differs from the previous "cold-cloud invigoration" concept (*13*), which does not consider aerosol impacts on condensational heating. As illustrated in Fig. 1, the enhanced condensational heating is driven mainly by the activation of $UAP_{<50}$ well above the cloud base, which is attributable to high vapor supersaturation with respect to water (S_w) resulting from fast droplet coalescence. The subsequent condensational growth considerably lowers the water supersaturation, liberating additional latent heating in the lower atmosphere, which enhances convective intensity drastically. The additional condensate from this process increases the amounts of both warm rain and supercooled cloud water. Furthermore, enhanced droplet freezing, including riming, occurs as a result of the production of additional supercooled water, and intensified convection enhances ice deposition at high levels.

[1]Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA. [2]Institute of Earth Sciences, Hebrew University of Jerusalem, Jerusalem 91904, Israel. [3]Department of Atmospheric and Oceanic Science and ESSIC, University of Maryland, College Park, MD 20742, USA. [4]Environmental and Climate Sciences Department, Brookhaven National Laboratory, Upton, NY, USA. [5]State Laboratory of Earth Surface Process and Resource Ecology, GCESS, Beijing Normal University, Beijing, China. [6]Centro de Previsão de Tempo e Estudos Climáticos, Instituto Nacional de Pesquisas Espaciais, Cachoeira Paulista, Brazil. [7]School of Engineering and Applied Sciences and Department of Earth and Planetary Sciences, Harvard University, Boston, MA, USA. [8]Beijing Municipal Weather Modification Office, Beijing 100089, China. [9]Instituto de Física, Universidade de São Paulo, Rua do Matão 1371, 05508-090 São Paulo, Brazil. [10]Department of Physics, University of Maryland Baltimore County, Baltimore, MD, USA. [11]State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing 100086, China. [12]Atmospheric and Climate Dynamics Laboratory, Institute of Atmospheric Sciences, Federal University of Alagoas, Brazil. [13]Multiphase Chemistry and Biogeochemistry Departments, Max Planck Institute for Chemistry, 55020 Mainz, Germany. [14]Johannes Gutenberg University Mainz, 55122 Mainz, Germany. [15]Department of Meteorology, School of Technology, Amazonas State University (UEA), Manaus-AM, Brazil.
*Corresponding author. Email: jiwen.fan@pnnl.gov

Cumulatively, these enhanced ice-related processes at middle and upper levels further enhance convection, but to a much lesser degree relative to enhanced condensation at low levels. The effects of $UAP_{<50}$ contrast with those of $CCN_{>50}$: $UAP_{<50}$ increase warm rain rate and have no effect on the timing of peak rain rate, whereas $CCN_{>50}$ suppress warm rain rate and delay the timing of peak rain rate.

Observed relationships of convective intensity and precipitation with aerosols

The measurements from the U.S. Department of Energy (DOE) Atmospheric Radiation Measurement 1290-MHz Ultra High Frequency (UHF) Radar Wind Profilers [RWP (*31*, *32*)] during GoAmazon 2014/5 provide the most accurate data on vertical air motions and precipitation rates for Amazon DCCs, allowing us to directly examine the relationship of updraft intensity to aerosols. We focused on convective cells of local origin (*33*, *34*), favoring relatively simple and similar dynamics across such DCCs in the wet season. The intrusion of the Manaus pollution plume provides a natural laboratory for exploring pollution aerosol effects. Our analysis period was the 2014 wet season (1 March to 31 May). Figure 2 shows the correlation of mean vertical velocity estimated for the top 10th percentile of updrafts in each convective event with the respective aerosol particle number concentration (N_a), averaged over a 30-min period before the start of convection at the T3 site (~70 km southwest of Manaus, downwind). The N_a at the T3 site varied substantially from case to case. We found that convective updraft velocity (w) increased with an increase of N_a for aerosols with a diameter (D) larger than 15 nm (Fig. 2A, top). When the updraft velocity was segregated according to four aerosol groups (ranging from lower to higher N_a), we observed a monotonic and striking increase of w, which increased from 4 m s^{-1} for the lowest-N_a group to 10 m s^{-1} for the highest-N_a group (Fig. 2B, left). The corresponding radar reflectivity (a measure of precipitation intensity) also increased with increasing N_a throughout the vertical profile for D > 15 nm (Fig. 2C, left), associated with hydrometeor size and precipitation rate increases. Remarkably, the increasing trend in updraft intensity and radar reflectivity as N_a increases does not hold well when considering only those aerosols with D > 50 nm (Fig. 2A, bottom, and Fig. 2B, right; similar results were obtained for D > 100 nm). The result suggests that $UAP_{<50}$, not $CCN_{>50}$, are the primary drivers for intensified convection. The probability density functions of w and rain rate from the four aerosol groups indicate that occurrences of stronger updraft velocities are more frequent with the increase of N_a for $UAP_{<50}$ (fig. S1A, left) and that the maximum rain rates also increase (fig. S2C).

To corroborate that $UAP_{<50}$ are the main factor contributing to the observed DCC enhancements, we conducted additional analyses to help isolate aerosol effects from thermodynamic controls. Locally driven Amazon deep convective events within the wet season should initiate and evolve under similar diurnal controls on their dynamical and thermodynamical environments (*33*, *35*). We examined traditional radiosonde thermodynamic forcing parameters such as the convective available potential energy (CAPE) and convective inhibition (CIN) before convection. The CAPE is very similar for these events (Fig. 2A, black circles) and is not strongly correlated with updraft intensity. Under similar CAPE conditions, CIN magnitudes should help to reveal how likely it is for DCCs to initiate and to determine what may be their relative intensities. CIN varies quite a bit (Fig. 2A, triangles) but again shows no correlation with the enhanced convective intensity as N_a of $UAP_{<50}$ increases. Analyses of profiles of the temperature, relative humidity (RH), and zonal U- and meridional V-components of the wind fields representative of the pre-storm environment also indicate that these environmental profiles do not correlate with an increase of updraft intensity as N_a of $UAP_{<50}$ increases (fig. S3). In fact, our lowest-N_a group exhibited higher RH at 2- to 5-km altitudes than did the higher-N_a groups, which should have favored stronger convection and offset some aerosol effects. This means that the trend for enhanced updraft intensity with the increase of N_a counting $UAP_{<50}$ should have potentially been more prominent if RH for the lowest-N_a group is similar to those of the higher-N_a groups. Although we cannot guarantee that RWP observations captured the representative updraft cores for every single event, these analyses do not show any covariation of aerosols with dynamics and thermodynamics for these locally occurring systems. This provides clear evidence that the enhanced convective intensity seen with the increase in N_a of $UAP_{<50}$ is not solely controlled by factors other than $UAP_{<50}$.

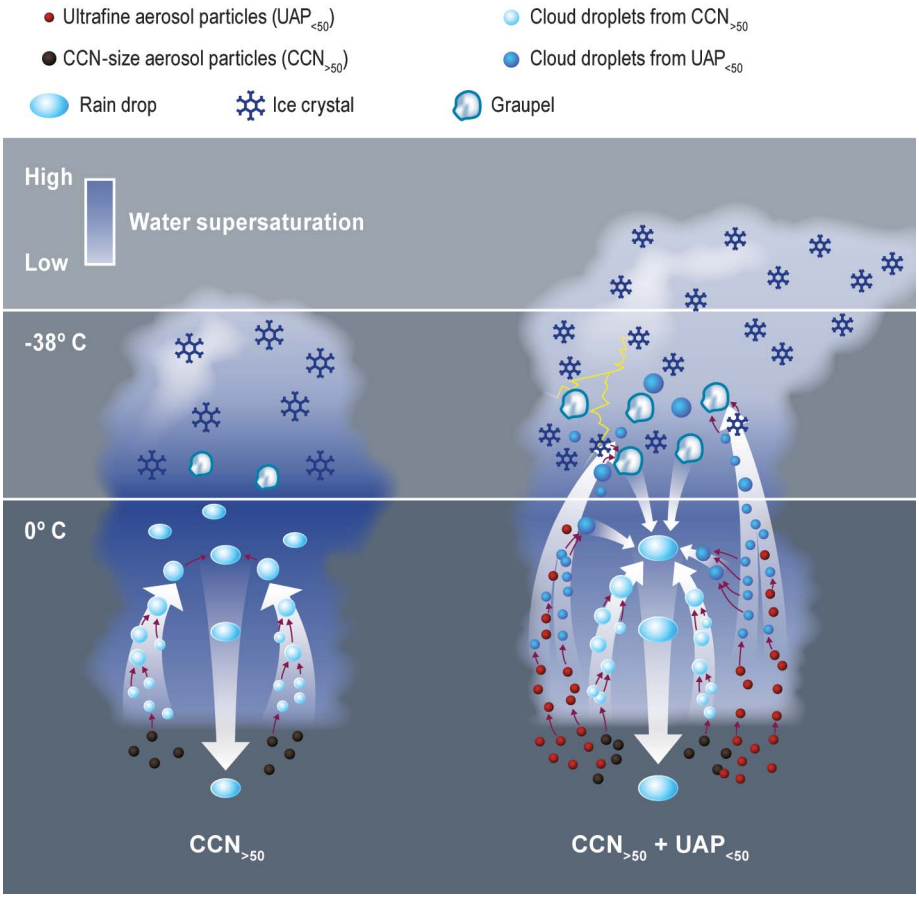


Fig. 1. Illustration of the effect of ultrafine aerosol particles ($UAP_{<50}$) on tropical convective clouds. In clouds that lack $UAP_{<50}$ (left), the clouds are highly supersaturated as a result of fast drop coalescence that forms warm rain and reduces the integrated droplet surface area available for condensation. With added $UAP_{<50}$ (right, red dots), an additional number of cloud droplets are nucleated above cloud base, which lowers supersaturation drastically by enhanced condensation, releasing additional latent heat at low and middle levels, thus intensifying convection. The additional condensate adds to both the warm rain and supercooled cloud water; when freezing occurs aloft, this addition further enhances convection (i.e., a small increase in convection but enhancement of precipitation and storm electrification).

Modeled mechanism

To understand the physical processes and mechanisms responsible for the observed intensification of updrafts by $UAP_{<50}$, we conducted model simulations at a cloud-resolving scale of 0.5 km for a typical wet season convective event, as on 17 March

2014 (fig. S4) (*34*). We used detailed spectral-bin microphysics (*36*, *37*) coupled with the Weather Research and Forecasting model (*38*). The simulation for the observed case ("baseline") is P3_BG, using a value of 950 cm^{-3} (130 $CCN_{>50}$ + 820 $UAP_{<50}$) for the present-day background aerosol concentration; the concentration in the Manaus metropolis (fig. S5B, black box) was set to be higher by a factor of 3 for both $CCN_{>50}$ and $UAP_{<50}$ to account for the effect of the Manaus pollution plume (Table 1). The aerosol size distribution (SD) is shown in fig. S5C. The power-law SD for $UAP_{<50}$ followed the observed shape in the city of Manaus (fig. S6C, line 1). For $CCN_{>50}$ (D > 50 nm), the SD is based on data from a remote rainforest site known as the Amazon Tall Tower Observatory (ATTO) in the central Amazon Basin (*39*, *40*), approximating a clean preindustrial (PI) condition (C_PI). C_BG is the simulation based on P3_BG, except with the Manaus pollution plume removed. To explore the effects of $UAP_{<50}$, we performed another pair of simulations, C_PI and PL3_PI, based on C_BG and P3_BG (Table 1), respectively, except with $UAP_{<50}$ removed (fig. S5C). Therefore, C_PI represents the PI condition without $UAP_{<50}$. The vertical distribution (VD) is shown in fig. S6A, and the details are in (*34*). Besides the four main simulations used to demonstrate the mechanism and isolate the contributions, we also conducted sensitivity tests (Table 1) with a different aerosol VD (C_BG_VD and C_PI_VD) and SD (P3_BG_SD, C_BG_SD, and C_PI_SD) to examine how the proposed mechanism is affected by variously measured aerosol properties. The sensitivity test C_PI2 was intended to show that the results do not change qualitatively even if a small amount of $UAP_{<50}$, as observed in the pristine condition over the Amazon, exists (*34*).

We evaluated the baseline simulation (P3_BG) with available observations (figs. S7 to S9) (*34*). Overall, the simulation is in good agreement with the observed case in terms of profiles of temperature, water vapor, and wind fields (fig. S7); the precipitation rate and the timing of peak precipitation (figs. S8 and S9A); and the echo-top heights of 10 dBZ (fig. S9B). These observational metrics provide confidence in the model performance.

With the sensitivity tests based on the baseline simulation, we found that adding $UAP_{<50}$ to the PI environment without $UAP_{<50}$ markedly invigorates convective intensity (Fig. 3A, C_PI versus C_BG and PL3_PI versus P3_BG). The probability density functions of updraft velocity clearly convey the same point (fig. S10). Therefore, the model replicates results that were observed: $UAP_{<50}$, which are abundant in the Manaus pollution plume, lead to enhanced convection around the T3 area. The convective intensity is more than 50% weaker in the PI environment (C_PI) when compared to the perturbed environment by the urban plume (P3_BG). As a result, the peak rain rate is greater by a factor of nearly 2.5 (Fig. 3C). Even under the present-day Manaus regional background condition that already contains an appreciable amount of $UAP_{<50}$ (820 cm^{-3}), further increasing $UAP_{<50}$ due to the Manaus pollution plume by a factor of 3 still drastically enhances convective intensity (by about 40%) in the downwind area (P3_BG versus C_BG). The corresponding increase in the peak precipitation rate is nearly 100%, much more substantial than that of convective intensity. This result highlights the ubiquity of convective invigoration by aerosols in Amazonia (i.e., in both PI and present-day Manaus regional background environments) and the nonlinear response of precipitation rates to convective intensity. The ubiquity of convective invigoration by the Manaus pollution plume is also corroborated by much more frequent strong updrafts over the entire domain from C_PI to C_BG

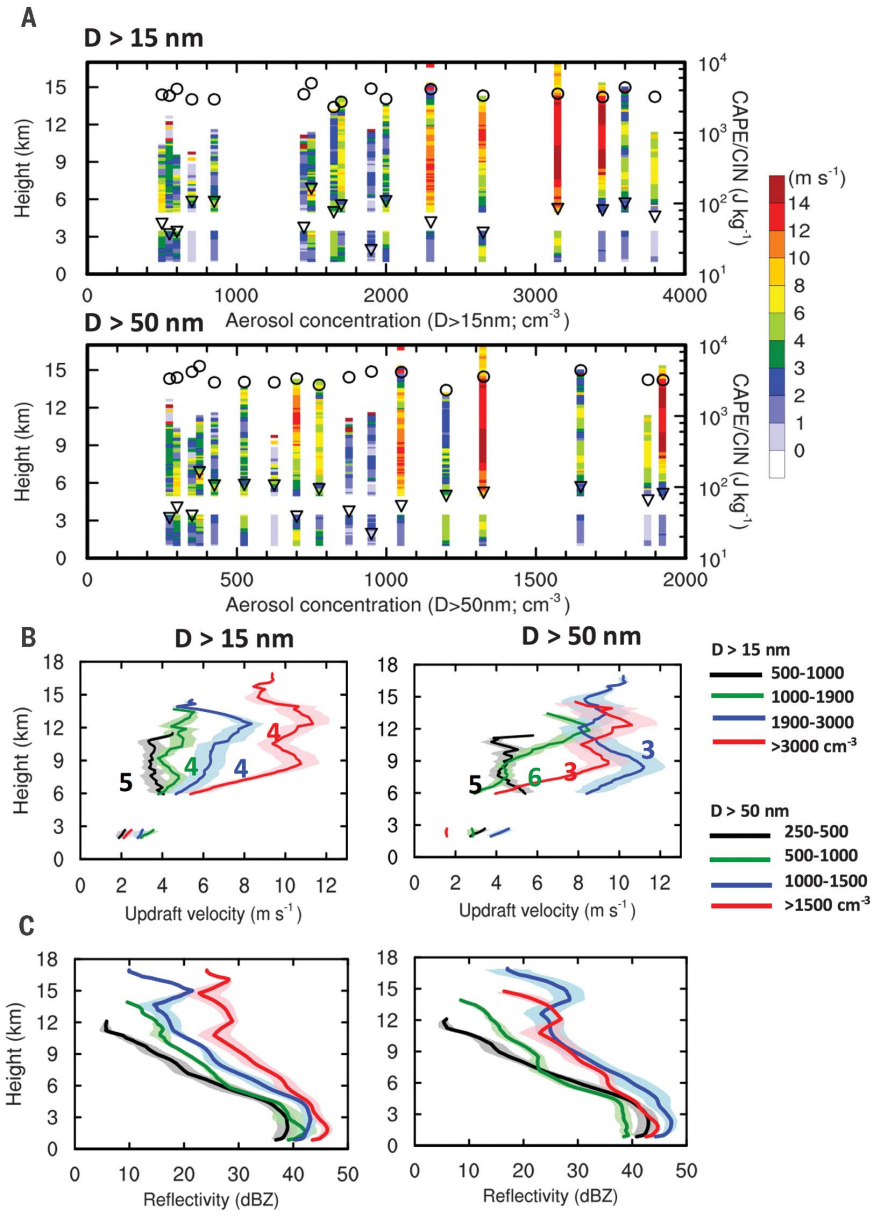


Fig. 2. Observed updraft velocity (w) and radar reflectivity (Z_e) as aerosol increases. (**A**) Vertical profiles of updraft velocity averaged over the 90th to 100th percentile of updrafts in each selected local convective case as a function of N_a with D > 15 nm (top) and D > 50 nm (bottom). Each profile column represents one case. The circles and triangles denote the CAPE and CIN values of each case, respectively. The aerosol value for each case is listed in table S1. (**B**) Vertical profiles of the updraft velocity averaged over the top 10 percentiles of updrafts for each case for the four aerosol groups from lower to higher N_a for D > 15 nm (left) and D > 50 nm (right). The number of cases for each group is marked. (**C**) Same as (B) except for reflectivity. Values of (B) and (C) are the means of the collection of the top 10 percentiles of updrafts from each case. The top 10 percentiles of the updrafts have 20 to 40 data points for most of the vertical levels and a few for some high levels in each case. The data have a vertical resolution of ~120 m and time resolution of ~6 s. Shaded areas in (B) and (C) represent the standard error of the data.

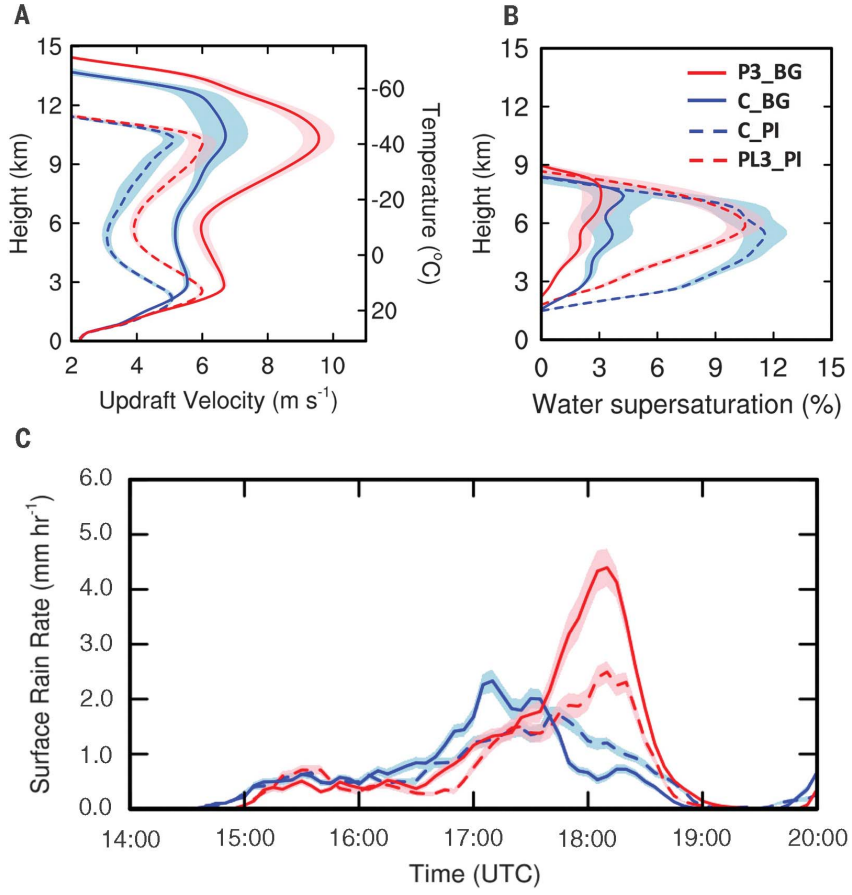


Fig. 3. Simulated aerosol effects on the DCCs. (**A** and **B**) Vertical profiles of updraft velocity *w* (A) and water supersaturation (B) averaged over the top 10 percentiles (i.e., 90th to 100th) for the updrafts with $w > 2$ m s^{-1} during 1400–1900 UTC from the convective clouds around the T3 site (red box in fig. S8). (**C**) Time series of mean surface rain rate averaged over the red box area from simulations of C_PI (blue dashed curve), C_BG (blue solid curve), PL3_PI (red dashed curve), and P3_BG (red solid curve). The right-side *y* axis in (A) shows the temperature profile. The convective clouds over T3 were chosen for analysis because they are affected by the Manaus pollution plume in P3_BG and evaluated by observations as shown in figs. S7 to S9. Comparisons with radar-retrieved rain rates at 2.5-km altitude are shown in figs. S8 and S9A. Shaded areas represent the standard error of the data.

(fig. S11). Another notable feature is that the timing of the peak rain rate is delayed when increasing $CCN_{>50}$, consistent with previous studies. However, adding $UAP_{<50}$ to the same CCN condition does not change the timing, although it greatly enhances the rain rate resulting from enhanced convective intensity. The reason for no change in the timing of rain by $UAP_{<50}$ is discussed later.

Further analysis shows that the mechanism responsible for such strongly enhanced convective intensity by $UAP_{<50}$ is different from the "cold-phase invigoration" previously proposed (*13*). We used comparisons between C_PI and C_BG to demonstrate our proposed mechanism (similar conclusions were drawn from comparisons between PL3_PI and P3_BG). The warm- and deep-cloud periods were examined. The vertical velocity increase under the deep-cloud period correlates with the increase in total buoyancy (no correlation with cold pool intensity is seen), although the condensate loading effect partially offsets the thermal buoyancy (Fig. 4B, bottom). The thermal buoyancy increase is attributed to the increased net latent heating (fig. S12A).

Next, we detail how latent heating is increased and which microphysical processes are the major contributors. In the PI environment (C_PI), for the warm-cloud period after warm rain initiates, the mean S_w for the top 10th percentile of updrafts can be nearly 4% (Fig. 4A, top) because of the efficient warm rain formation due to fast droplet coalescence resulting from large droplet size, which reduces the integrated droplet surface area for condensation (Fig. 5A, top). The addition of $UAP_{<50}$ (C_BG) brings the S_w down to 1%, because $UAP_{<50}$ form an additional number of droplets above the cloud base—a result of smaller critical aerosol size for activation under high S_w (Fig. 4A, second from top). These droplets increase the integrated surface area above the cloud base by more than 40% (Fig. 5A, top) and therefore promote condensational growth even at the warm-cloud period (Fig. 4A, third from top). Note that enhanced condensation leads not only to the increased cloud water but also to a rainwater increase under the warm-cloud period (Fig. 5A, bottom). This result is corroborated by aircraft measurements of warm clouds during GoAmazon2014/5, showing larger liquid water content in the plume-affected clouds relative to background clouds; enhanced condensation efficiency is part of the explanation (*41*). Under the deep-cloud period, the condensation effect is further amplified; the S_w reaches up to 15% in C_PI because of a lack of droplet surface area for condensation and decreases to 7% in C_BG (Figs. 4B and 5B, top). This allows aerosols as small as 10 nm to be activated into cloud droplets (Fig. 4B, second from top). The activation of $UAP_{<50}$ leads to a factor of 2 to 4 increase of the integrated droplet surface area and a doubled condensational heating rate (Fig. 4B, third from top). The decrease of supersaturation by $UAP_{<50}$ is similarly evident even over the entire convective period 1400–1900 UTC (Fig. 3B).

Because of the very small critical size for droplet nucleation at 2.5- to 3-km altitude, another droplet nucleation peak appears above the peak at the cloud base in C_BG; this is attributable to the presence of a large amount of $UAP_{<50}$ (Fig. 4B, top, pink solid line). Without $UAP_{<50}$, cloud droplet nucleation occurs predominantly at the cloud base because almost all of the aerosol particles can be activated there (Fig. 4B, top, pink dashed line). The addition of $UAP_{<50}$ makes droplet nucleation occur at a large rate above the cloud base and up to an altitude of 8 km as a result of transport of $UAP_{<50}$ from low levels. The enhanced droplet nucleation over the vertical profile from C_PI to C_BG leads to an increase of up to a factor of 2 to 3 in condensational heating occurring throughout the lower and middle troposphere. The magnitude of condensational heating in C_BG is nearly twice that of depositional heating, indicating the predominance of the condensation process in the release of latent heat. The increase in depositional heating from C_PI to C_BG is also much less than that of condensational heating and is only located at 10- to 12-km altitude, whereas condensational heating is located at low and middle levels and has a deep layer of increase. The latent heating rates

from riming and drop freezing increase considerably from C_PI to C_BG as well. The increases in drop freezing and riming rates result from the transport of additional droplets from low levels. Although the increasing rate in latent heating from drop freezing is the largest, the actual magnitude is too small (relative to the other processes) for latent heat to play an important role. The latent heat from riming is also substantially smaller than that from condensation and deposition.

To understand the relative role of the enhanced condensational heating and latent heating from ice-related processes in intensifying convection, we conducted three sensitivity tests based on P3_BG (*34*). These tests show that enhanced condensational heating is the predominant player in the intensification of convection and precipitation, while enhanced latent heating at upper levels plays a secondary role (figs. S13 and S14). This is different from "cold-phase invigoration" (i.e., enhanced ice-related processes), which is caused by the addition of CCN that can be activated at the cloud base via suppressing coalescence and delaying warm rain. $UAP_{<50}$ do not suppress the initial droplet coalescence and delay the rain, because $UAP_{<50}$ can only be activated above the cloud base when rain has already formed and supersaturation is enhanced by a reduced droplet surface area available for condensation. Although previous idealized simulations suggested that smaller aerosol particles can enhance DCC intensity through condensational heating (*14*), our observations and simulations show that it actually occurs and has surprisingly large magnitude and importance.

Because of this greatly enhanced convection associated with $UAP_{<50}$, the mass mixing ratios of all hydrometeors are nearly doubled under the deep-cloud period (Fig. 5B, bottom). The activation of $UAP_{<50}$ leads to an increase—up to an order of magnitude—in droplet number concentrations (Fig. 5B, middle), as well as a drastic increase in ice-phase particle concentrations resulting from enhanced freezing including riming. All ice-phase particle number concentrations (ice, snow, and graupel) increase by a factor of ~2 generally, but the increase of the mass mixing ratio of ice-phase particles is largely driven by graupel, with snow mass reduced (fig. S12B, left), particularly in mixed-phase regions (6 to 10 km), resulting from a larger amount of supercooled liquid droplets in the cases of $UAP_{<50}$. Falling graupel particles contribute to warm-phase increases in rain mass under the deep-cloud period (Fig. 5B, bottom), thereby contributing to increased surface precipitation as in Fig. 3C.

Summary and discussion

The retrieved updraft velocity from RWP has allowed us to directly examine and constrain aerosol impacts on updraft intensity for DCCs occurring in a similar dynamic and thermodynamic environment except for aerosols, revealing the potentially substantial convective updraft and precipitation enhancements by $UAP_{<50}$ from the Manaus pollution plume. We used real-case three-dimensional simulations to reproduce observed convective cloud characteristics and to observe convective updraft and precipitation enhancements by $UAP_{<50}$ over the Amazon region. The physical mechanism for such strong intensification of convection stems from the strong capacity of these DCCs in activating $UAP_{<50}$ that usually have a much higher number concentration than $CCN_{>50}$, because fast droplet coalescence in pristine conditions decreases integrated droplet surface area for condensational growth, producing highly supersaturated conditions. The subsequent condensational growth of an additional number of droplets considerably lowers the water supersaturation, liberating a large amount of additional latent heat at the low and middle levels of DCCs and considerably enhancing updraft strength. This "warm-phase invigoration" has much stronger

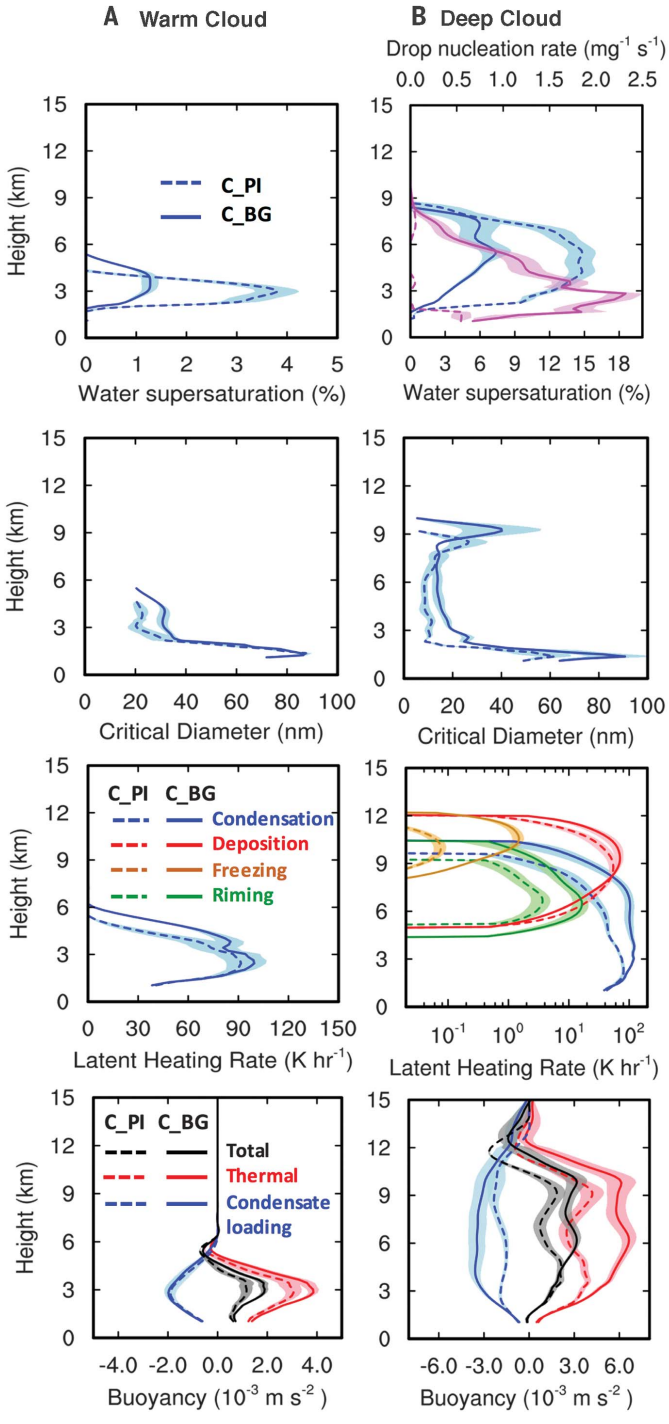


Fig. 4. Analysis of thermodynamics and microphysics. (**A** and **B**) Vertical profiles for the warm cloud (A) and deep cloud (B) from C_PI (dashed curves) and C_BG (solid curves). From top to bottom: water supersaturation; critical diameter for activation; latent heating from condensation (blue), deposition (red), drop freezing (orange), and riming (green); and buoyancy terms. The dashed and solid pink curves in the top panel of (B) show droplet nucleation rates from C_PI and C_BG, respectively (upper *x* axis). The thermal buoyancy includes contributions from both temperature and water vapor variations. The values for the warm cloud are averaged over the top 10 percentiles (i.e., 90th to 100th) of the updrafts with $w > 1$ m s^{-1} from a 30-min duration after the warm rain starts and the rain rate exceeds 0.5 mm hour^{-1} for the convective clouds in the red box in fig. S8. The values for the deep cloud are averaged over the top 10 percentiles (i.e., 90th to 100th) of the updrafts with $w > 2$ m s^{-1} from a 30-min duration with 15 min before and after the strongest convection. Therefore, the specific time for the warm- and deep-cloud periods varies by simulation because of the time shift of convective clouds between the simulations. Shaded areas represent the standard error of the data.

effects than the "cold-phase invigoration" previously proposed (*13*). $UAP_{<50}$ increase the rainwater amount mainly through enhanced accretion of added cloud droplets and added graupel melting. Added $UAP_{<50}$ do not affect the timing of precipitation because $UAP_{<50}$ can be activated to form additional cloud droplets only after warm rain begins. In contrast, $CCN_{>50}$ suppress and delay warm rain and then delay peak precipitation.

The large magnitude of the condensation effects is likely a result of the very warm, moist, and clean conditions in the Amazon wet season. This leads to a mechanism different from that associated with dry-season events over the Amazon [i.e., cold-phase invigoration (*13*, *22*–*24*)] or other dryer or colder locations [e.g., (*11*, *42*, *43*)]. A key factor is the ample cloud droplet coalescence in the tropical rainforest environment that is not affected by human activities. This coalescence rapidly removes droplets; droplet removal leads to suppressed condensation, leaving a high supersaturation for activating $UAP_{<50}$. This is corroborated by sensitivity tests with droplet collision-coalescence processes shut off for both background (C_BG) and plume-affected (P3_BG) cases, which show small differences in water supersaturation between the two cases, in contrast to the large differences when droplet collision-coalescence processes occur.

It has been believed that there is a cutoff size—typically larger than 60 nm (*20*)—for aerosol particles to be able to serve as CCN for a specific region or typical aerosol type (*18*). Generally, $UAP_{<50}$ are not considered to contribute to CCN because their small sizes require high supersaturation. Our findings show that under the present-day Manaus regional background, approximated PI, and present-day pristine environments over the Amazon region, $UAP_{<50}$ can effectively serve as CCN for droplet nucleation well above the cloud base after coalescence starts. DCCs in the natural rainforest environment over the Amazon, where fine particles are scarce, may be especially susceptible to the addition of $UAP_{<50}$, as shown by this work. The dearth of new particle formation in the Amazon boundary layer under natural conditions might be a result of low sulfuric acid concentration; another possible reason is that the products of gas-phase oxidation of isoprene do not facilitate new particle formation as efficiently as those from oxidation of other biogenic volatile organic compounds (e.g., monoterpenes). However, anthropogenic emissions in this region may induce aerosol nucleation and produce a large amount of $UAP_{<50}$ through various mechanisms (*44*, *45*); for instance, interactions of anthropogenic sulfur and NO_x with biogenic sources such as organic acids from aromatics enhance aerosol nucleation.

A recent study showed that aerosol concentrations over the Amazon peak at the upper troposphere (*46*), unlike the VD assumed in our model simulations. We conducted sensitivity tests (*34*) by using a VD similar to that in figure 7B of (*46*). Similar convective invigoration by increasing $UAP_{<50}$ through the same mechanism is seen (fig. S15, A and B). The new VD does not affect the conclusion because the maximum supersaturation peaks at an altitude of 6 km and the latent heating at low levels is the major contributor to convection intensification, so the addition of $UAP_{<50}$ at upper levels would not greatly affect the convective invigoration below it. However, the new VD causes stronger convective invigoration at upper levels because of higher aerosol concentrations. Another uncertainty results from the aerosol SD assumed in the simulation. We also carried out sensitivity tests by using a SD measured by a fast integrated mobility spectrometer [FIMS (*47*)] on aircraft at ~20 km downwind from Manaus (*34*). The results were similar to the previous simulations with a different SD (fig. S15, C and D). Thus, we infer that any aerosol scenario that contains a large ratio of $UAP_{<50}$ to $CCN_{>50}$ in a humid environment would produce the $UAP_{<50}$ effects we report in this study.

Because all $UAP_{<50}$ are removed in our simulations for examining the $UAP_{<50}$ effects, and because small numbers of $UAP_{<50}$ are observed at the remote rainforest site in the central Amazon Basin [e.g., figure 6A of (*39*)], we conducted an additional sensitivity test, C_PI2, in which $UAP_{<50}$ (60 cm^{-3}) were added to C_PI (Table 1) (*34*). We found that the differences in updraft velocity and supersaturation between C_PI and C_PI2 are very small (fig. S16, A and B). Thus, our conclusion and the proposed mechanism are not affected by the small amount of $UAP_{<50}$ in the PI condition. This indicates that we should expect similarly large $UAP_{<50}$ effects for pristine regions with low $UAP_{<50}$ concentration. The intrusion of urban areas such as the Manaus metropolis produces a large amount of $UAP_{<50}$, which may drastically change convective and precipitation conditions over the Amazon Basin by producing stronger convective clouds and precipitation.

All of our tests with different aerosol properties (SD and VD) indicate that observed and simulated $UAP_{<50}$ effects and the proposed mechanism are robust. Given the importance of the Amazon Basin in global circulation and hydrology, the ultrafine pollution particles introduced by human activities could have profound effects on other places around the globe. A lack of aerosol nucleation (or a low concentration of ultrafine particles) and the existence of ample moisture may not be unique to the Amazon Basin and may also occur in other tropical forests and oceans, further highlighting the global importance of our results. For example, a recent study found a conspicuous enhancement of lightning over shipping lanes in the equatorial Indian Ocean (*48*). We conclude that $UAP_{<50}$ can modulate convection and rain-forming processes more strongly than previously considered, especially in pristine regions. Therefore, the changes

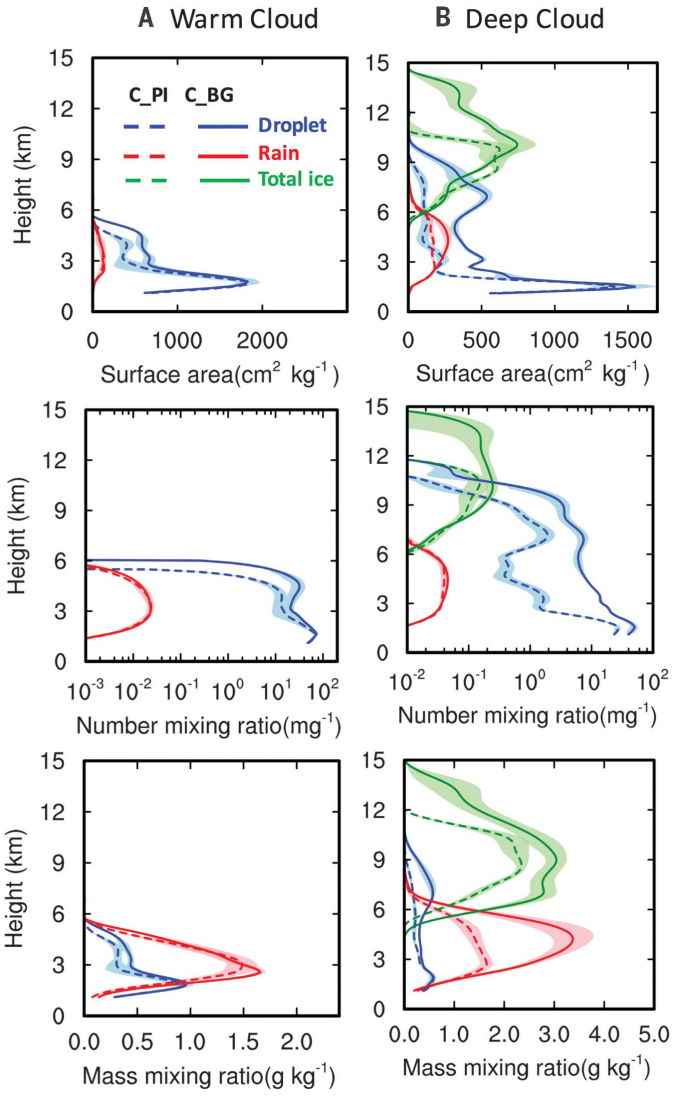


Fig. 5. Additional microphysical properties. (**A** and **B**) Analogous to Fig. 4, showing vertical profiles for droplet surface area (top), droplet number (middle), and mass mixing ratios (bottom) for cloud droplets (blue), rain drops (red), and ice-phase particles (green). The ice-phase particles include cloud ice, snow, and graupel. Shaded areas represent the standard error of the data.

Table 1. Model simulations. Abbreviations: BG, present-day background aerosol condition in the Manaus region; PI, preindustrial aerosol condition; P3, factor of 3 increase in N_a for plume; C, clean condition (i.e., no plume); L in PL3, large aerosol particles (i.e., D > 50 nm). Peaked refers to a measured size distribution that has a peak value over a size range; upper-level peak refers to a measured vertical distribution with aerosols peaked at upper levels (*34*). N/A means that no plume is applied (i.e., aerosols are horizontally uniform over the domain at the model initial time).

Simulation	Acronym	Size distribution	Vertical distribution	Hygroscopicity (κ)	N_a of domain (per cubic centimeter)		N_a of Manaus pollution plume (per cubic centimeter)	
					$CCN_{>50}$	$UAP_{<50}$	$CCN_{>50}$	$UAP_{<50}$
Manaus regional background with pollution plume	P3_BG	Power law + peaked	Exponential decrease	0.12	130	820	390	2460
Manaus regional background	C_BG	Power law + peaked	Exponential decrease	0.12	130	820	N/A	N/A
PI condition without $UAP_{<50}$	C_PI	Peaked	Exponential decrease	0.12	130	0	N/A	N/A
Sensitivity test for P3_BG without $UAP_{<50}$	PL3_PI	Peaked	Exponential decrease	0.12	130	0	390	0
Sensitivity test for VD based on C_BG	C_BG_VD	Power law + peaked	Upper-level peak	0.12	130	820	N/A	N/A
Sensitivity test for VD based on C_PI	C_PI_VD	Peaked	Upper-level peak	0.12	130	0	N/A	N/A
Sensitivity test for SD based on P3_BG	P3_BG_SD	Peaked	Exponential decrease	0.12	450	3350	1350	10,050
Sensitivity test for SD based on C_BG	C_BG_SD	Peaked	Exponential decrease	0.12	450	3350	N/A	N/A
Sensitivity test for SD based on C_PI	C_PI_SD	Peaked	Exponential decrease	0.12	450	0	N/A	N/A
PI condition with $UAP_{<50}$	C_PI2	Peaked	Exponential decrease	0.12	130	60	N/A	N/A
Sensitivity test for C_PI with mean ATTO	C_PI_ATTO	Peaked	Exponential decrease	0.12	200	0	N/A	N/A
Sensitivity test for C_BG with $CCN_{>50}$ from mean ATTO	C_BG_ATTO	Peaked	Exponential decrease	0.12	200	820	N/A	N/A

of ultrafine pollution particles from preindustrial conditions to the present may possibly have appreciably changed deep convective clouds.

REFERENCES AND NOTES

1. R. A. Houze Jr., *Cloud Dynamics* (Elsevier/Academic Press, ed. 2, 2014).
2. H. C. Barnes, R. A. Houze Jr., *J. Geophys. Res. Atmos.* **121**, 8269–8296 (2016).
3. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2013).
4. J. Fan, Y. Wang, D. Rosenfeld, X. Liu, *J. Atmos. Sci.* **73**, 4221–4252 (2016).
5. W. K. Tao, J. P. Chen, Z. Li, C. Wang, C. Zhang, *Rev. Geophys.* **50**, RG2001 (2012).
6. A. P. Khain, *Environ. Res. Lett.* **4**, 015004 (2009).
7. S. S. Lee, L. J. Donner, V. T. J. Phillips, Y. Ming, *J. Geophys. Res.* **113**, D16202 (2008).
8. Z. J. Lebo, J. H. Seinfeld, *Atmos. Chem. Phys.* **11**, 5407–5429 (2011).
9. R. L. Storer, S. C. van den Heever, *J. Atmos. Sci.* **70**, 430–446 (2013).
10. S. C. van den Heever, G. L. Stephens, N. B. Wood, *J. Atmos. Sci.* **68**, 699–718 (2011).
11. Z. Li *et al.*, *Nat. Geosci.* **4**, 888–894 (2011).
12. X. Yang, Z. Li, *J. Geophys. Res. Atmos.* **119**, 1835–1844 (2014).
13. D. Rosenfeld *et al.*, *Science* **321**, 1309–1313 (2008).
14. A. P. Khain, V. Phillips, N. Benmoshe, A. Pokrovsky, *J. Atmos. Sci.* **69**, 2787–2807 (2012).
15. Q. Chen *et al.*, *Atmos. Chem. Phys.* **17**, 9585–9598 (2017).
16. J. Fan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, E4581–E4590 (2013).
17. U. Pöschl *et al.*, *Science* **329**, 1513–1516 (2010).
18. M. O. Andreae, *Science* **315**, 50–51 (2007).
19. S. T. Martin *et al.*, *Atmos. Chem. Phys.* **10**, 11415–11438 (2010).
20. U. Lohmann, J. Feichter, *Atmos. Chem. Phys.* **5**, 715–737 (2005).
21. I. Koren, G. Dagan, O. Altaratz, *Science* **344**, 1143–1146 (2014).
22. M. O. Andreae *et al.*, *Science* **303**, 1337–1342 (2004).
23. I. Koren, J. V. Martins, L. A. Remer, H. Afargan, *Science* **321**, 946–949 (2008).
24. J. C. Lin, T. Matsui, R. A. Pielke Sr., C. Kummerow, *J. Geophys. Res.* **111**, D19204 (2006).
25. J. C. Zhou, E. Swietlicki, H. C. Hansson, P. Artaxo, *J. Geophys. Res.* **107**, 8055 (2002).
26. M. Kulmala *et al.*, *Atmos. Chem. Phys.* **11**, 13061–13143 (2011).
27. J. Wang *et al.*, *Nature* **539**, 416–419 (2016).
28. M. Pikridas *et al.*, *Atmos. Chem. Phys.* **15**, 10219–10237 (2015).
29. S. T. Martin *et al.*, *Bull. Am. Meteorol. Soc.* **98**, 981–997 (2017).
30. S. T. Martin *et al.*, *Atmos. Chem. Phys.* **16**, 4785–4797 (2016).
31. S. E. Giangrande *et al.*, *J. Appl. Meteorol. Climatol.* **52**, 2278–2295 (2013).
32. S. E. Giangrande *et al.*, *J. Geophys. Res. Atmos.* **121**, 12891–12913 (2016).
33. S. Tang *et al.*, *Atmos. Chem. Phys.* **16**, 14249–14264 (2016).
34. See supplementary materials.
35. J. A. Marengo *et al.*, *Atmos. Chem. Phys. Discuss.* 10.5194/acp-2017-22 (2017).
36. A. P. Khain, A. Pokrovsky, M. Pinsky, A. Seifert, V. Phillips, *J. Atmos. Sci.* **61**, 2963–2982 (2004).
37. J. Fan *et al.*, *J. Geophys. Res.* **117**, D00K36 (2012).
38. W. C. Skamarock *et al.*, *A Description of the Advanced Research WRF Version 3* (National Center for Atmospheric Research, 2008).
39. M. L. Pöhlker *et al.*, *Atmos. Chem. Phys. Discuss.* 10.5194/acp-2017-847 (2017).
40. M. L. Pöhlker *et al.*, *Atmos. Chem. Phys.* **16**, 15709–15740 (2016).
41. M. A. Cecchini *et al.*, *Atmos. Chem. Phys.* **16**, 7029–7041 (2016).
42. R. Zhang, G. Li, J. Fan, D. L. Wu, M. J. Molina, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5295–5299 (2007).
43. Y. Wang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 6894–6899 (2014).
44. R. Zhang, A. Khalizov, L. Wang, M. Hu, W. Xu, *Chem. Rev.* **112**, 1957–2011 (2012).
45. M. Shrivastava *et al.*, *Rev. Geophys.* **55**, 509–559 (2017).
46. M. O. Andreae *et al.*, *Atmos. Chem. Phys. Discuss.* 10.5194/acp-2017-694 (2017).
47. J. Wang, M. Pikridas, S. R. Spielman, T. Pinterich, *J. Aerosol Sci.* **108**, 44–55 (2017).
48. J. A. Thornton, K. S. Virts, R. H. Holzworth, T. P. Mitchell, *Geophys. Res. Lett.* **44**, 9102–9111 (2017).

ACKNOWLEDGMENTS

This study was supported by the U.S. DOE, Office of Science, Atmospheric System Research Program. The Pacific Northwest National Laboratory (PNNL) is operated for DOE by Battelle Memorial Institute under contract DE-AC06-76RLO1830. This research used PNNL Institutional Computing resources. Y.Z. and Z.L. were supported by NSF grant AGS1534670 and National Science Foundation of China grant 91544217. D.R. was supported by project BACCHUS European Commission FP7-603445. S.E.G. represents Brookhaven Science Associates LLC under DOE contract DE-SC0012704. The DOE Atmospheric Radiation Measurement (ARM) Climate Research Facility's

GoAmazon field campaign data were used. The x-band and s-band (SIPAM) radar data were supported by the CHUVA project. We thank the GoAmazon team and the CHUVA team for their effort to produce the observational data. We acknowledge support from the Central Office of the Large Scale Biosphere Atmosphere Experiment in Amazonia (LBA), Instituto Nacional de Pesquisas da Amazonia (INPA), Universidade do Estado do Amazonas (UEA), and the local Research Foundation (FAPEAM). L.A.T.M., P.A., and H.M.J.B. were supported by FAPESP grants 2009/15235-8, 2013/05014-0, and 2013/50510-5. The work was conducted under authorization 001030/2012-4 of the Brazilian National Council for Scientific and Technological Development (CNPq). For the operation of the ATTO site, we acknowledge support by the German Federal Ministry of Education and Research (BMBF contract 01LB1001A), the Brazilian Ministério da Ciência, Tecnologia e Inovação (MCTI/FINEP contract 01.11.01248.00), and the Amazon State University (UEA), FAPEAM, LBA/INPA, and SDS/CEUC/RDS-Uatumã. We thank C. Schumacher and A. Funk at Texas A&M University for the SIPAM data, T. Biscaro for the x-band data, S. Tang at Lawrence Livermore National Laboratory for the input of local convective system selection, S. Hagos at PNNL for the input of model configuration, and C. Kuang at Brookhaven National Laboratory for help in understanding the uncertainty of Scanning Mobility Particle Sizer data. The observational data including x-band and SIPAM radar data from CHUVA can be obtained from DOE ARM data archive www.archive.arm.gov/discovery/#v/results/s/fiop::amf2014goamazon, which is available to the community. The model simulation data are archived at PNNL PIC and are available at https://dtn2.pnl.gov/data/jiwen/GoAmazon_simulations_sci/.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6374/411/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S17
Table S1
References (*49*–*53*)

26 May 2017; accepted 22 December 2017
10.1126/science.aan8461

Cite as: N. Brown, T. Sandholm, *Science*
10.1126/science.aao1733 (2017).

Superhuman AI for heads-up no-limit poker: Libratus beats top professionals

Noam Brown and Tuomas Sandholm*

Computer Science Department, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA.

***Corresponding author. Email: sandholm@cs.cmu.edu**

No-limit Texas hold'em is the most popular form of poker. Despite AI successes in perfect-information games, the private information and massive game tree have made no-limit poker difficult to tackle. We present Libratus, an AI that, in a 120,000-hand competition, defeated four top human specialist professionals in heads-up no-limit Texas hold'em, the leading benchmark and long-standing challenge problem in imperfect-information game solving. Our game-theoretic approach features application-independent techniques: an algorithm for computing a blueprint for the overall strategy, an algorithm that fleshes out the details of the strategy for subgames that are reached during play, and a self-improver algorithm that fixes potential weaknesses that opponents have identified in the blueprint strategy.

In recent years the field of artificial intelligence (AI) has advanced considerably. The measure of this progress has, in many cases, been marked by performance against humans in benchmark games. AI programs have defeated top humans in checkers (*1*), chess (*2*), and Go (*3*). In these perfect-information games both players know the exact state of the game at every point. In contrast, in imperfect-information games, some information about the state of the game is hidden from a player—for example, the opponent may hold hidden cards. Hidden information is ubiquitous in real-world strategic interactions, such as business strategy, negotiation, strategic pricing, finance, cybersecurity, and military applications, which makes research on general-purpose techniques for imperfect-information games particularly important.

Hidden information makes a game far more complex for a number of reasons. Rather than simply search for an optimal sequence of actions, an AI for imperfect-information games must determine how to balance actions appropriately, so that the opponent never finds out too much about the private information the AI has. For example, bluffing is a necessary feature in any competitive poker strategy, but bluffing all the time would be a bad strategy. In other words, the value of an action depends on the probability it is played.

Another key challenge is that different parts of the game cannot be considered in isolation; the optimal strategy for a given situation may depend on the strategy that would be played in situations that have not occurred (*4*). As a consequence, a competitive AI must always consider the strategy for the game as a whole.

Poker has a long history as a challenge problem for developing AIs that can address hidden information (*5*–*11*). No-limit Texas hold'em is the most popular form of poker in the world. The heads-up (that is, two-player) variant prevents opponent collusion and kingmaker scenarios where a bad player causes a mediocre player to shine, and therefore allows a clear winner to be determined. Due to its large size and strategic complexity, heads-up no-limit Texas hold'em (HUNL) has been the primary benchmark and challenge problem for imperfect-information game solving for several years. No prior AI has defeated top human players in this game.

In this paper we introduce Libratus, (*12*) an AI that takes a distinct approach to addressing imperfect-information games. In a 20-day, 120,000-hand competition featuring a $200,000 prize pool, it defeated top human professionals in HUNL. The techniques in Libratus do not use expert domain knowledge or human data and are not specific to poker; thus they apply to a host of imperfect-information games.

Game-solving approach in Libratus

Libratus features three main modules:

(i) The first module computes an abstraction of the game, which is smaller and easier to solve, and then computes game-theoretic strategies for the abstraction. The solution to this abstraction provides a detailed strategy for the early rounds of the game, but only an approximation for how to play in the more numerous later parts of the game. We refer to the solution of the abstraction as the blueprint strategy.

(ii) When a later part of the game is reached during play, the second module of Libratus constructs a finer-grained abstraction for that subgame and solves it in real time (*13*). Unlike subgame-solving techniques in perfect-information games, Libratus does not solve the subgame abstraction in isolation; instead, it ensures that the fine-grained solution to

the subgame fits within the larger blueprint strategy of the whole game. The subgame solver has several key advantages over prior subgame-solving techniques (*14*, *15*, *16*). Whenever the opponent makes a move that is not in the abstraction, a subgame is solved with that action included. We call this nested subgame solving. This technique comes with a provable safety guarantee.

(iii) The third module of Libratus—the self-improver—enhances the blueprint strategy. It fills in missing branches in the blueprint abstraction and computes a game-theoretic strategy for those branches. In principle, one could conduct all such computations in advance, but the game tree is way too large for that to be feasible. To tame this complexity, Libratus uses the opponents' actual moves to suggest where in the game tree such filling is worthwhile.

In the following three subsections, we present these three modules in more detail.

Abstraction and equilibrium finding: Building a blueprint strategy

One solution to the problem of imperfect information is to simply reason about the entire game as a whole, rather than just pieces of it. In this approach, a solution is pre-computed for the entire game, possibly using a linear program (*10*) or an iterative algorithm (*17*–*21*). For example, an iterative algorithm called counterfactual regret minimization plus (CFR+) was used to near-optimally solve heads-up limit Texas hold'em, a relatively simple version of poker, which has about 10^{13} unique decision points (*11*, *22*).

In contrast, HUNL (*23*) has 10^{161} decision points (*24*), so traversing the entire game tree even once is impossible. Precomputing a strategy for every decision point is infeasible for such a large game.

Fortunately, many of those decision points are very similar. For example, there is little difference between a bet of $100 and a bet of $101. Rather than consider every possible bet between $100 and $20,000, we could instead just consider increments of $100. This is referred to as action abstraction. An abstraction is a smaller, simplified game that retains as much as possible the strategic aspects of the original game. This drastically reduces the complexity of solving the game. If an opponent bets $101 during an actual match, then the AI may simply round this to a bet of $100 and respond accordingly (*25*–*27*). Most of the bet sizes included in Libratus's action abstraction were nice fractions or multiples of the pot [roughly determined by analyzing the most common bet sizes at various points in the game taken by prior top AIs in the Annual Computer Poker Competition (ACPC) (*28*)]. However, certain bet sizes early in the game tree were determined by an application-independent parameter optimization algorithm that converged to a locally optimal set of bet sizes (*29*).

An additional form of abstraction is abstraction of actions taken by chance, that is, card abstraction in the case of poker. Similar hands are grouped together and treated identically. Intuitively, there is little difference between a King-high flush and a Queen-high flush. Treating those hands as identical reduces the complexity of the game and thus makes it computationally easier. Nevertheless, there are still differences even between a King-high flush and a Queen-high flush. At the highest levels of play, those distinctions may be the difference between winning and losing. Libratus does not use any card abstraction on the first and second betting rounds. The last two betting rounds, which have a significantly larger number of states, are abstracted only in the blueprint strategy. The 55 million different hand possibilities on the third round were algorithmically grouped into 2.5 million abstract buckets, and the 2.4 billion different possibilities on the fourth round were algorithmically grouped into 1.25 million abstract buckets. However, the AI does not follow the blueprint strategy in these rounds and instead applies nested subgame solving, described in the next section, which does not use any card abstraction. Thus, each poker hand is considered individually during actual play. The card abstraction algorithm that we used was similar to that used in our prior AIs Baby Tartanian8 (*30*), which won the 2016 ACPC, and Tartanian7 (*31*–*33*), which won the 2014 ACPC (there was no ACPC in 2015).

Once the abstraction was constructed, we computed the blueprint strategy for Libratus by having the AI play simulated games of poker against itself (while still exploring the hypothetical outcomes of actions not chosen) using an improved version of an algorithm called Monte Carlo Counterfactual Regret Minimization (MCCFR). MCCFR (*17*, *34*, *35*) has a long history of use in successful poker AIs (*30*, *31*, *36*, *37*). MCCFR maintains a regret value for each action. Intuitively, regret represents how much the AI regrets having not chosen that action in the past. When a decision point is encountered during self play, the AI chooses actions with higher regret with higher probability (*38*). As more and more games are simulated, MCCFR guarantees that with high probability a player's average regret for any action (total regret divided by the number of iterations played) approaches zero. Thus, the AI's average strategy over all simulated games gradually improves. We will now describe the equilibrium-finding algorithm (*4*).

On each simulated game, MCCFR chooses one player (who we refer to as the traverser) that will explore every possible action and update his regrets, while the opponent simply plays according to the strategy determined by the current regrets. The algorithm switches the roles of the two players after each game, that is, a single hand of poker. Every time either player is faced with a decision point in a simulated game, the player will choose a probability distribution over actions based on regrets on those actions (which are determined by what he had learned in earlier games when he had

been in that situation). For the first game, the AI has not learned anything yet and therefore uses a uniform random distribution over actions. At traverser decision points, MCCFR explores every action in a depth-first manner. At opponent decision points, MCCFR samples an action based on the probability distribution. This process repeats at every decision point until the game is over and a reward is received, which is passed up. When a reward is returned by every action at a traverser decision point, MCCFR calculates the weighted average reward for that decision point based on the probability distribution over actions. The regret for each action is then updated by adding the value returned by that action, and subtracting the weighted average reward for the decision point. The weighted average reward is then passed up to the preceding decision point, and so on.

Our improved version of MCCFR traverses a smaller portion of the game tree on each iteration. Intuitively, there are many clearly suboptimal actions in the game, and repeatedly exploring them wastes computational resources that could be better used to improve the strategy elsewhere. Rather than explore every hypothetical alternative action to see what its reward would have been, our algorithm probabilistically skips over unpromising actions that have very negative regret as it proceeds deeper into the tree during a game (*30*, *39*). This led to a factor of three speedup of MCCFR in practice and allowed us to solve larger abstractions than were otherwise possible.

This skipping also mitigates the problems that stem from imperfect recall. The state-of-the-art practical abstractions in the field, including ours, are imperfect-recall abstractions where some aspects of the cards on the path of play so far are intentionally forgotten in order to be able to computationally afford to have a more detailed abstraction of the present state of cards (*30*–*32*, *40*). Since all decisions points in a single abstract card bucket share the same strategy, updating the strategy for one of them leads to updating the strategy for all of them. This is not an issue if all of them share the same optimal strategy at the solution reached, but in practice there are differences between their optimal strategies and they effectively "fight" to push the bucket's strategy toward their own optimal strategy. Skipping negative-regret actions means that decision points that will never be reached in actual play will no longer have their strategies updated, thereby allowing the decision points that will actually occur during play to move the bucket's strategy closer to their optimal strategies.

We ran our algorithm on an abstraction that is very detailed in the first two rounds of HUNL, but relatively coarse in the final two rounds. However, Libratus never plays according to the abstraction solution in the final two rounds. Rather, it uses the abstract blueprint strategy in those rounds only to estimate what reward a player should expect to receive with a particular hand in a subgame. This estimate is used to determine a more precise strategy during actual play, as described in the next section.

Nested safe subgame solving

Although purely abstraction-based approaches have produced strong AIs for poker (*25*, *30*, *32*, *41*), abstraction alone has not been enough to reach superhuman performance in HUNL. In addition to abstraction, Libratus builds upon prior research into subgame solving (*14*–*16*, *42*), in which a more detailed strategy is calculated for a particular part of the game that is reached during play. Libratus features many advances in subgame solving that proved critical to achieving superhuman performance (*43*).

Libratus plays according to the abstract blueprint strategy only in the early parts of HUNL, where the number of possible states is relatively small and we can afford the abstraction to be extremely detailed. Upon reaching the third betting round, or any earlier point in the game where the remaining game tree is sufficiently small (*44*), Libratus constructs a new, more detailed abstraction for the remaining subgame and solves it in real time.

However, there is a major challenge with subgame solving in imperfect-information games: a subgame cannot be solved in isolation because its optimal strategy may depend on other, unreached subgames (*4*). Prior AIs that used real-time subgame solving addressed this problem by assuming the opponent plays according to the blueprint strategy. However, the opponent can exploit this assumption by simply switching to a different strategy. For this reason, the technique may produce strategies that are far worse than the blueprint strategy and is referred to as unsafe subgame solving (*42*, *45*). Safe subgame solving techniques, on the other hand, guarantee that the subgame's new strategy makes the opponent no better off no matter what strategy the opponent might use (*14*). They accomplish this by ensuring that the new strategy for the subgame fits within the overarching blueprint strategy of the original abstraction. Ensuring the opponent is no better off relative to the blueprint strategy is trivially possible because we could just reuse the blueprint strategy. However, now that the abstraction is more detailed in the subgame and we can better distinguish the strategic nuances of the subgame, it may be possible to find an improvement over the prior strategy that makes the opponent worse off no matter what cards she is holding.

We now describe Libratus's core technique for determining an improved strategy in a subgame. For exposition, we assume Player 2 (P2) is determining an improved strategy against Player 1 (P1). Given that P2's strategy outside the subgame is σ_2, there exists some optimal strategy σ_2^* that P2

could play in the subgame. We would like to find or approximate σ_2^* in real time. We assume that, for each poker hand P1 might have, we have a good estimate of the value P1 receives in the subgame with that hand by playing optimally against σ_2^*, even though we do not know σ_2^* itself. Although we do not know these values exactly, we can approximate them with the values P1 receives in the subgame in the blueprint strategy. We later prove that if these estimates are approximately accurate, we can closely approximate σ_2^*.

To find a strategy close to σ_2^* in the subgame using only the values from the blueprint, we create an augmented subgame (Fig. 1) which contains the subgame and additional structures. At the start of the augmented subgame, P1 is privately dealt a random poker hand. Given that P2 plays according to σ_2 prior to the subgame, and given P1's dealt hand, there is a particular probability distribution over what hands P2 might have in this situation. P2 is dealt a poker hand according to this probability distribution. P1 then has the choice of entering the subgame (which is now far more detailed than in the blueprint strategy), or of taking an alternative payoff that ends the augmented subgame immediately. The value of the alternative payoff is our estimate, according to the blueprint strategy, of P1's value for that poker hand in that subgame. If P1 chooses to enter the subgame, then play proceeds normally until the end of the game is reached. We can solve this augmented subgame just as we did for the blueprint strategy (*46*).

For any hand P1 might have, P1 can do no worse in the augmented subgame than just choosing the alternative payoff (which awards our estimate of the expected value P1 could receive against σ_2^*). At the same time, P2 can ensure that for every poker hand P1 might have, he does no better than what he could receive against σ_2^*, because P2 can simply play σ_2^* itself. Thus, any solution to the augmented subgame must do approximately as well as σ_2^*—where the approximation error depends on how far off our estimates of P1's values are. P2 then uses the solution to the augmented subgame as P2's strategy going forward.

All of this relied on the assumption that we have accurate estimates of P1's values against σ_2^*. Although we do not know these values exactly, we can approximate them with values from the blueprint strategy. We now prove that if these estimates are approximately accurate, subgame solving will produce a strategy that is close to the quality of σ_2^*. Specifically, we define the exploitability of a strategy σ_2 as how much more σ_2 would lose, in expectation, against a worst-case opponent than what P2 would lose, in expectation, in an exact solution of the full game.

Theorem 1 uses a form of safe subgame solving we coin Estimated-Maxmargin. We define a margin for every P1 hand in a subgame as the expected value of that hand according to the blueprint minus what P1 could earn with that hand, in expectation, by entering the more-detailed subgame. Estimated-Maxmargin finds a strategy that maximizes the minimum margin among all P1 hands. It is similar to a prior technique called Maxmargin (*15*) except that the prior technique conservatively used as the margin what P1 could earn in the subgame, in expectation, by playing a best response to P2's blueprint strategy minus what P1 could earn, in expectation, by entering the more-detailed subgame.

Theorem 1. *Let σ_i be a strategy for a two-player zero-sum perfect-recall game, let S be a set of non-overlapping subgames in the game, and let σ_i^* be the least-exploitable strategy that differs from σ_i only in S. Assume that for any opponent decision point (hand in the case of poker) and any subgame in S, our estimate of the opponent's value in a best response to σ_i^* for that decision point in that subgame is off by at most Δ. Applying Estimated-Maxmargin subgame solving to any subgame in S reached during play results in overall exploitability at most 2Δ higher than that of σ_i^* (47).*

Although safe subgame solving techniques have been known for three years (*14*, *15*), they were not used in practice because empirically they performed significantly worse than unsafe subgame solving (*42*) head to head (*48*). Libratus features a number of advances to subgame solving that greatly improve effectiveness.

(i) Although we describe safe subgame solving as using estimates of P1 values, past techniques used upper bounds on those values (*14*, *15*). Using upper bounds guarantees that the subgame solution has exploitability no higher than the blueprint strategy. However, it tends to lead to overly conservative strategies in practice. Using estimates can, in theory, result in strategies with higher exploitability than the blueprint strategy, but Theorem 1 bounds how much higher this exploitability can be.

(ii) It arrives at better strategies in subgames than was previously thought possible. Past techniques ensured that the new strategy for the subgame made P1 no better off in that subgame for every situation. It turns out that this is an unnecessarily strong constraint. For example, 2♠7♥ is considered the worst hand in HUNL and should be folded immediately, which ends the game. Choosing any other action would result in an even bigger loss in expectation. Nevertheless, past subgame solving techniques would be concerned about P1 having 2♠7♥ in a subgame, which is unrealistic. Even if subgame solving resulted in a strategy that increased the value of 2♠7♥ a small amount in one subgame, that increase would not outweigh the cost of reaching the

subgame (that is, the cost of not folding with 2♠7♥). Thus, P2 can allow the value of some "unimportant" P1 hands to increase in subgames, so long as the increase is small enough that it is still a mistake for P1 to reach the subgame with that hand. We accomplish this by increasing the alternative reward of P1 hands in the augmented subgame by the extra cost to P1 of reaching the subgame, that is, the size of the mistake P1 would have to make to reach that subgame with that hand. By increasing the alternative reward in the augmented subgame of these "unimportant" hands, P2 develops a strategy in the subgame that better defends against hands P1 might actually have (*4*).

(iii) Libratus crafts a unique strategy in response to opponent bets, rather than rounding it to the nearest size in the abstraction. The optimal response to a bet of $101 is different from the optimal response to a bet of $100, but the difference is likely minor. For that reason, rounding an opponent bet of $101 to $100 is reasonable. But the optimal response to a bet of $150 is likely significantly different from the response to a bet of $100 or a bet of $200. In principle one could simply increase the number of actions in the abstraction, perhaps by considering bets in increments of $10 rather than $100, so that the error from rounding is smaller. However, the size of the abstraction, and the time needed to solve it, increases prohibitively as more actions are added.

Therefore, rather than round to the nearest action, Libratus calculates a unique response in real time to off-tree actions, that is, an action taken by an opponent that is not in the abstraction. Libratus attempts to make the opponent no better off, no matter what hand the opponent might have, for having chosen the off-tree action rather than an action in the abstraction. It does this by generating and solving an augmented subgame following the off-tree action where the alternative payoff is the best in-abstraction action the opponent could have taken (the best action may differ across hands).

Libratus repeats this for every subsequent off-tree action in a process we call nested subgame solving (see Fig. 2). Later we provide experiments that demonstrate that this technique improves the worst-case performance of poker AIs by more than an order of magnitude compared to the best technique for rounding opponent actions to a nearby in-abstraction action.

(iv) Because the subgame is solved in real time, the abstraction in the subgame can also be decided in real time and change between hands. Libratus leverages this feature by changing, at the first point of subgame solving, the bet sizes it will use in that subgame and every subsequent subgame of that poker hand, thereby forcing the opponent to continually adapt to new bet sizes and strategies (*49*).

The authors of the poker AI DeepStack independently and concurrently developed an algorithm similar to nested subgame solving, which they call continual re-solving (*50*). In an Internet experiment, DeepStack defeated human professionals who are not specialists in HUNL. However, DeepStack was never shown to outperform prior publicly-available top AIs in head-to-head performance, whereas Libratus beats the prior leading HUNL poker AI Baby Tartanian8 by a wide margin, as we discuss later.

Like Libratus, DeepStack computes in real time a response to the opponent's specific bet and uses estimates rather than upper bounds on the opponent's values. It does not share Libratus's improvement of de-emphasizing hands the opponent would only be holding if she had made an earlier mistake, and does not share the feature of changing the subgame action abstraction between hands.

DeepStack solves a depth-limited subgame on the first two betting rounds by estimating values at the depth limit via a neural network. This allows it to always calculate real-time responses to opponent off-tree actions, while Libratus typically plays according to its pre-computed blueprint strategy in the first two rounds.

Because Libratus typically plays according to a pre-computed blueprint strategy on the first two betting rounds, it rounds an off-tree opponent bet size to a nearby in-abstraction action. The blueprint action abstraction on those rounds is dense in order to mitigate this weakness. In addition, Libratus has a unique self-improvement module to augment the blueprint strategy over time, which we now introduce.

Self-improvement

The third module of Libratus is the self-improver. It enhances the blueprint strategy in the background. It fills in missing branches in the blueprint abstraction and computes a game-theoretic strategy for those branches. In principle, one could conduct all such computations in advance, but the game tree is way too large for that to be feasible. To tame this complexity, Libratus uses the opponents' actual moves to suggest where in the game tree such filling is worthwhile.

The way machine learning has typically been used in game playing is to try to build an opponent model, find mistakes in the opponent's strategy (e.g., folding too often, calling too often, etc.), and exploit those mistakes (*51–53*). The downside is that trying to exploit the opponent opens oneself to being exploited. (A certain conservative family of exploitation techniques constitutes the sole exception to this downside (*51–53*).) For that reason, to a first approximation, Libratus did not do opponent exploitation. Instead, it used the data of the bet sizes that the opponents used to suggest which branches should be added to the blueprint, and it then computed game-theoretic strategies for those branches in the background.

In most situations that can occur in the first two betting rounds, real-time subgame solving as used in Libratus would likely not produce a better strategy than the blueprint, because the blueprint already uses no card abstraction in those rounds and conducting subgame solving in real time so early in the game tree would require heavy abstraction in the subgame. For these reasons, Libratus plays according to the precomputed blueprint strategy in these situations. In those rounds there are many bet sizes in the abstraction, so the error from rounding to a nearby size is small. Still, there is some error, and this could be reduced by including more bet sizes in the abstraction. In the experiment against human players described in the next section, Libratus analyzed the bet sizes in the first betting round most heavily used by its opponents in aggregate during each day of the competition. Based on the frequency of the opponent bet sizes and their distance from the closest bet size in the abstraction, Libratus chose k bet sizes for which it would try to calculate a response overnight (*54*). Each of those bet sizes for which reasonable convergence had been reached by the morning was then added to the blueprint strategy together with the newly-computed strategy following that bet size. In this way Libratus was able to progressively narrow its gaps as the competition proceeded by leveraging the humans' ability to find potential weaknesses. Furthermore, these fixes to its strategy are universal: they work against all opponents, not just the opponents that Libratus has faced.

Libratus's self-improvement comes in two forms. For one of them, when adding one of the k bet sizes, a default sibling bet size is also used during the equilibrium finding so as to not assume that the opponent necessarily only uses the bet size that will be added. For the other, a default bet size is not used. This can be viewed as more risky and even exploitative, but Libratus mitigates the risk by using that part of the strategy during play only if the opponent indeed uses that bet size most of the time (*4*).

Experimental evaluation

To evaluate the strength of the techniques used in Libratus, we first tested the overall approach of the AI on scaled-down variants of poker before proceeding to tests on full HUNL. These moderate-sized variants consisted of only two or three rounds of betting rather than four, and at most three bet sizes at each decision point. The smaller size of the games allowed us to precisely calculate exploitability, the distance from an optimal strategy. Performance was measured in milli-big blinds per hand (mbb/hand), the average number of big blinds won per 1,000 hands.

In the first experiment, we compared using no subgame solving, unsafe subgame solving (*42*) (in which a subgame is solved in isolation with no theoretical guarantees on performance), and safe subgame solving just once upon reaching the final betting round of the game. Both players were constrained to choosing among only two different bet sizes, so off-tree actions were not an issue in this first experiment. The results are shown in Table 1. In all cases, safe subgame solving reduced exploitability by more than a factor of 4 relative to no subgame solving. In one case, unsafe subgame solving led to even lower exploitability, while in another it increased exploitability by nearly an order of magnitude more than if no subgame solving had been used. This demonstrates that although unsafe subgame solving may produce strong strategies in some games, it may also lead to far worse performance. Safe subgame solving, in contrast, reduced exploitability in all games.

In the second experiment, we constructed an abstraction of a game which only includes two of the three available bet sizes. If the opponent played the missing bet size, the AI either used action translation [in which the bet is rounded to a nearby size in the abstraction; we compared against the leading action translation technique (*27*)], or nested subgame solving. The results are shown in Table 2. Nested subgame solving reduced exploitability by more than an order of magnitude relative to action translation.

Next we present experiments in full HUNL. After constructing Libratus, we tested the AI against the prior leading HUNL poker AI, our 2016 bot Baby Tartanian8, which had defeated all other poker AIs with statistical significance in the most recent ACPC (*55*). We report average win rates followed by the 95% confidence interval. Using only the raw blueprint strategy, Libratus lost to Baby Tartanian8 by 8 ± 15 mbb/hand. Adding state-of-the-art post-processing on the 3rd and 4th betting rounds (*31*), such as eliminating low-probability actions that are likely only positive owing to insufficient time to reach convergence, led to the Libratus blueprint strategy defeating Baby Tartanian8 by 18 ± 21 mbb/hand. Eliminating low-probability actions empirically leads to better performance against non-adjusting AIs. However, it also increases the exploitability of the AI because its strategy becomes more predictable. The full Libratus agent did not use post-processing on the third and fourth betting rounds. On the first two rounds, Libratus primarily used a new, more robust, form of post-processing (*4*).

The next experiment evaluated nested subgame solving (with no post-processing) using only actions that are in Baby Tartanian8's action abstraction. Libratus won by 59 ± 28 mbb/hand (*56*). Finally, applying the nested subgame solving structure used in the competition resulted in Libratus defeating Baby Tartanian8 by 63 ± 28 mbb/hand. The results are shown in Table 3. In comparison, Baby Tartanian8 defeated the next two strongest AIs in the ACPC by 12 ± 10 mbb/hand and 24 ± 20 mbb/hand.

Finally, we tested Libratus against top humans. In January 2017, Libratus played against a team of four top HUNL

specialist professionals in a 120,000-hand Brains vs. AI challenge match over 20 days. The participants were Jason Les, Dong Kim, Daniel McCauley, and Jimmy Chou. A prize pool of $200,000 was allocated to the four humans in aggregate. Each human was guaranteed $20,000 of that pool. The remaining $120,000 was divided among them based on how much better the human did against Libratus than the worst-performing of the four humans. Libratus decisively defeated the humans by a margin of 147 mbb/hand, with 99.98% statistical significance and a p-value of 0.0002 (if the hands are treated as independent and identically distributed), see Fig. 3 (*57*). It also beat each of the humans individually.

Conclusions

Libratus presents an approach that effectively addresses the challenge of game-theoretic reasoning under hidden information in a large state space. The techniques that we developed are largely domain independent and can thus be applied to other strategic imperfect-information interactions, including non-recreational applications. Owing to the ubiquity of hidden information in real-world strategic interactions, we believe the paradigm introduced in Libratus will be important for the future growth and widespread application of AI.

REFERENCES AND NOTES

1. J. Schaeffer, *One Jump Ahead: Challenging Human Supremacy in Checkers* (Springer, 1997).
2. M. Campbell, A. J. Hoane Jr., F.-H. Hsu, Deep Blue. *Artif. Intell.* **134**, 57–83 (2002). doi:10.1016/S0004-3702(01)00129-1
3. D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, D. Hassabis, Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016). doi:10.1038/nature16961 Medline
4. See supplementary materials for more details.
5. J. Nash, "Non-cooperative games," thesis, Princeton University (1950).
6. J. F. Nash, L. S. Shapley, *Contributions to the Theory of Games*, H. W. Kuhn, A. W. Tucker, Eds. (Princeton Univ. Press, 1950), vol. 1, pp. 105–116.
7. D. A. Waterman, Generalization learning techniques for automating the learning of heuristics. *Artif. Intell.* **1**, 121–170 (1970). doi:10.1016/0004-3702(70)90004-4
8. J. Shi, M. Littman, in *CG '00: Revised Papers from the Second International Conference on Computers and Games* (Springer, 2002), pp. 333–345.
9. D. Billings *et al*., in *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI)* (Morgan Kaufmann Publishers, San Francisco, 2003), pp. 661–668.
10. A. Gilpin, T. Sandholm, in *Proceedings of the National Conference on Artificial Intelligence (AAAI)* (AAAI Press, 2005), pp. 1684–1685.
11. M. Bowling, N. Burch, M. Johanson, O. Tammelin, Heads-up limit hold'em poker is solved. *Science* **347**, 145–149 (2015). doi:10.1126/science.1259433 Medline
12. Libratus is Latin and means balanced (for approximating Nash equilibrium) and forceful (for its powerful play style and strength).
13. An imperfect-information subgame (which we refer to simply as a subgame) is defined differently than how a subgame is usually defined in game theory. The usual definition requires that a subgame starts with the players knowing the exact state of the game, that is, no information is hidden from any player. Here, an imperfect-information subgame is determined by information that is common knowledge to the players. For example, in poker, a subgame is defined by the sequence of visible board cards and actions the players have taken so far. Every possible combination of private cards—that is, every node in the game tree which is consistent with the common knowledge—is a root of this subgame. Any node that descends from a root node is also included in the subgame. A formal definition is provided in the supplementary material.
14. N. Burch, M. Johanson, M. Bowling, in *AAAI Conference on Artificial Intelligence (AAAI)* (AAAI Press, 2014), pp. 602–608.
15. M. Moravcik, M. Schmid, K. Ha, M. Hladik, S. Gaukrodger, in *AAAI Conference on Artificial Intelligence (AAAI)* (AAAI Press, 2016).
16. E. Jackson, in *AAAI Workshop on Computer Poker and Imperfect Information* (AAAI Press, 2014).
17. M. Zinkevich, M. Johanson, M. H. Bowling, C. Piccione, in *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)* (Neural Information Processing Systems Foundation, Inc., 2007), pp. 1729–1736.
18. Y. Nesterov, Excessive gap technique in nonsmooth convex minimization. *SIAM J. Optim.* **16**, 235–249 (2005). doi:10.1137/S1052623403422285
19. S. Hoda, A. Gilpin, J. Peña, T. Sandholm, Smoothing techniques for computing Nash equilibria of sequential games. *Math. Oper. Res.* **35**, 494–512 (2010). doi:10.1287/moor.1100.0452
20. A. Gilpin, J. Peña, T. Sandholm, First-order algorithm with O(ln(1/ϵ)) convergence for ϵ-equilibrium in two-person zero-sum games. *Math. Program.* **133**, 279–298 (2012). doi:10.1007/s10107-010-0430-2.
21. C. Kroer, K. Waugh, F. Klnç-Karzan, T. Sandholm, in *Proceedings of the ACM Conference on Economics and Computation (EC)* (ACM, New York, 2017).
22. O. Tammelin, N. Burch, M. Johanson, M. Bowling, in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)* (AAAI Press, 2015), pp. 645–652.
23. The version of HUNL that we refer to, which is used in the Annual Computer Poker Competition, allows bets in increments of $1, with each player having $20,000 at the beginning of a hand.
24. M. Johanson, "Measuring the size of large no-limit poker games," (Technical Report, Univ. of Alberta Libraries, 2013).
25. A. Gilpin, T. Sandholm, T. B. Sørensen, in *Proceedings of the Seventh International Joint Conference on Autonomous Agents and Multiagent Systems* (International Foundation for Autonomous Agents and Multiagent Systems, 2008), vol. 2, pp. 911–918.
26. D. Schnizlein, M. Bowling, D. Szafron, in *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence* (AAAI Press, 2009), pp. 278–284.
27. S. Ganzfried, T. Sandholm, in *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence* (AAAI Press, 2013), pp. 120–128.
28. Annual Computer Poker Competition; www.computerpokercompetition.org.
29. N. Brown, T. Sandholm, in *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI)* (AAAI Press, 2014), pp. 594–601.
30. N. Brown, T. Sandholm, in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)* (AAAI Press, 2016), pp. 4238–4239.
31. N. Brown, S. Ganzfried, T. Sandholm, in *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems* (International Foundation for Autonomous Agents and Multiagent Systems, 2015), pp. 7–15.
32. N. Brown, S. Ganzfried, T. Sandholm, in *AAAI Conference on Artificial Intelligence (AAAI)* (AAAI Press, 2015), pp. 4270–4271.
33. M. Johanson, N. Burch, R. Valenzano, M. Bowling, in *Proceedings of the 2013 International Conference on Autonomous Agents and Multiagent Systems* (International Foundation for Autonomous Agents and Multiagent Systems, 2013), pp. 271–278.
34. M. Lanctot, K. Waugh, M. Zinkevich, M. Bowling, in *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)* (Neural Information Processing Systems Foundation, Inc., 2009), pp. 1078–1086.
35. R. Gibson, M. Lanctot, N. Burch, D. Szafron, M. Bowling, in *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence* (AAAI Press, 2012), pp. 1355–1361.
36. M. Johanson, N. Bard, M. Lanctot, R. Gibson, M. Bowling, in *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems* (International Foundation for Autonomous Agents and Multiagent Systems, 2012), vol. 2, pp. 837–846.

37. R. Gibson, "Regret minimization in games and the development of champion multiplayer computer poker-playing agents," thesis, University of Alberta (2014).
38. There are a number of theoretically correct ways to choose actions on the basis of their regrets. The most common is regret matching, in which an action is chosen in proportion to its positive regret (*58*). Another common choice is hedge (*59*, *60*).
39. An action *a* with regret $R(a)$ that is below a threshold C (where C is negative) is sampled with probability $K/[K + C - R(a)]$, where K is a positive constant. There is additionally a floor on the sample probability. This sampling is only done for about the last half of iterations to be run; the first half is conducted using traditional external-sampling MCCFR. Other formulas can also be used.
40. K. Waugh *et al*., in *Symposium on Abstraction, Reformulation, and Approximation (SARA)* (AAAI Press, 2009).
41. M. Johanson, N. Bard, N. Burch, M. Bowling, in *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence* (AAAI Press, 2012), pp. 1371–1379.
42. S. Ganzfried, T. Sandholm, in *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (International Foundation for Autonomous Agents and Multiagent Systems, 2015), pp. 37–45.
43. N. Brown, T. Sandholm, *Adv. Neural Inf. Process. Syst.* **30**, 689–699 (2017).
44. In Libratus, we considered "sufficiently small" to be situations where no additional bets or raises could be made.
45. Despite lacking theoretical guarantees, unsafe subgame solving empirically performs well in certain situations and requires less information to be precomputed. For this reason, Libratus uses it once upon first reaching the third betting round, while using safe subgame solving in all subsequent situations.
46. We solved augmented subgames using a heavily optimized form of the CFR+ algorithm (*22*, *61*) because of the better performance of CFR+ in small games where a precise solution is desired. The optimizations we use keep track of all possible P1 hands rather than dealing out a single one at random.
47. Note that the theorem only assumes perfect recall in the actual game, not in the abstraction that is used for computing a blueprint strategy. Furthermore, applying Estimated-Maxmargin assumes that that subroutine maximizes the minimum margin; a sufficient condition for doing so is that there is no abstraction in the subgame.
48. Indeed, the original purpose of safe subgame solving was merely to reduce space usage by reconstructing subgame strategies rather than storing them.
49. Specifically, Libratus increased or decreased all its bet sizes by a percentage chosen uniformly at random between 0 and 8%.
50. M. Moravčík, M. Schmid, N. Burch, V. Lisý, D. Morrill, N. Bard, T. Davis, K. Waugh, M. Johanson, M. Bowling, DeepStack: Expert-level artificial intelligence in heads-up no-limit poker. *Science* **356**, 508–513 (2017). doi:10.1126/science.aam6960 Medline
51. D. Billings, D. Papp, J. Schaeffer, D. Szafron, in *Proceedings of the National Conference on Artificial Intelligence (AAAI)* (AAAI Press, 1998), pp. 493–499.
52. S. Ganzfried, T. Sandholm, in *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (International Foundation for Autonomous Agents and Multiagent Systems, 2011).
53. S. Ganzfried, T. Sandholm, Safe opponent exploitation. *ACM Transaction on Economics and Computation (TEAC)* **3**, 1–28 (2015). doi:10.1145/2716322
54. Based on the available computing resources, we chose $k = 3$ so that the algorithm could typically fix three holes to reasonable accuracy in 24 hours.
55. Baby Tartanian8 and all other AIs in the ACPC are available to ACPC participants for benchmarking.
56. Baby Tartanian8 uses action translation in response to bet sizes that are not in its action abstraction. Our experiments above demonstrated that action translation performs poorly compared to subgame solving. Using only bet sizes in Baby Tartanian8's abstraction disentangles the effects of action translation from the improvement of nested subgame solving. Baby Tartanian8 still used actions that were not in Libratus's abstraction, and therefore the experiments can be considered conservative.
57. Because both the humans and the AI adapted over the course of the competition, treating the hands as independent is not entirely inappropriate. We include confidence figures to provide some intuition for the variance in HUNL. In any case, 147 mbb/hand over 120,000 hands is considered a massive and unambiguous victory in HUNL.
58. S. Hart, A. Mas-Colell, A simple adaptive procedure leading to correlated equilibrium. *Econometrica* **68**, 1127–1150 (2000). doi:10.1111/1468-0262.00153
59. N. Littlestone, M. K. Warmuth, The weighted majority algorithm. *Inf. Comput.* **108**, 212–261 (1994). doi:10.1006/inco.1994.1009
60. Y. Freund, R. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **55**, 119–139 (1997). doi:10.1006/jcss.1997.1504
61. M. Johanson, K. Waugh, M. Bowling, M. Zinkevich, in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)* (AAAI Press, 2011), pp. 258–265.
62. L. Kocsis, C. Szepesvári, in *European Conference on Maching Learning (ECML)* (Springer, 2006), pp. 282–293.
63. R. Coulom, *Computers and Games* (Springer, 2007), pp. 72–83.
64. D. E. Knuth, R. W. Moore, An analysis of alpha-beta pruning. *Artif. Intell.* **6**, 293–326 (1975). doi:10.1016/0004-3702(75)90019-3
65. J. F. Nash, Equilibrium points in *n*-person games. *Proc. Natl. Acad. Sci. U.S.A.* **36**, 48–49 (1950). doi:10.1073/pnas.36.1.48 Medline
66. N. Brown, T. Sandholm, in *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)* (2015), pp. 1972–1980.
67. N. Brown, T. Sandholm, in *International Conference on Machine Learning* (Proceedings of Machine Learning Research, 2017).
68. S. Ganzfried, T. Sandholm, K. Waugh, in *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (International Foundation for Autonomous Agents and Multiagent Systems, 2012), pp. 871–878.

ACKNOWLEDGMENTS

This material is based on research supported by the National Science Foundation under grants IIS-1718457, IIS-1617590, and CCF-1733556, and by the ARO under award W911NF-17-1-0082, as well as XSEDE computing resources provided by the Pittsburgh Supercomputing Center. The Brains vs. AI competition was sponsored by Carnegie Mellon University, Rivers Casino, GreatPoint Ventures, Avenue4Analytics, TNG Technology Consulting, Artificial Intelligence, Intel, and Optimized Markets, Inc. We thank Ben Clayman for computing statistics of the play of our AIs against humans. The data presented in this paper are shown in the main text and supplementary material. Additional data can be obtained from the corresponding author upon request. Because HUNL poker is played commercially, the risk associated with releasing the code outweighs the benefits. To aid reproducibility, we have included the pseudo-code for the major components of our program in (*4*). The technology has been exclusively licensed to Strategic Machine, Inc., and the authors have ownership interest in the company.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/cgi/content/full/science.aao1733/DC1
Supplementary text
Figs. S1 and S2
Table S1
References (*62*–*68*)

22 June 2017; accepted 12 December 2017
Published online 17 December 2017
10.1126/science.aao1733

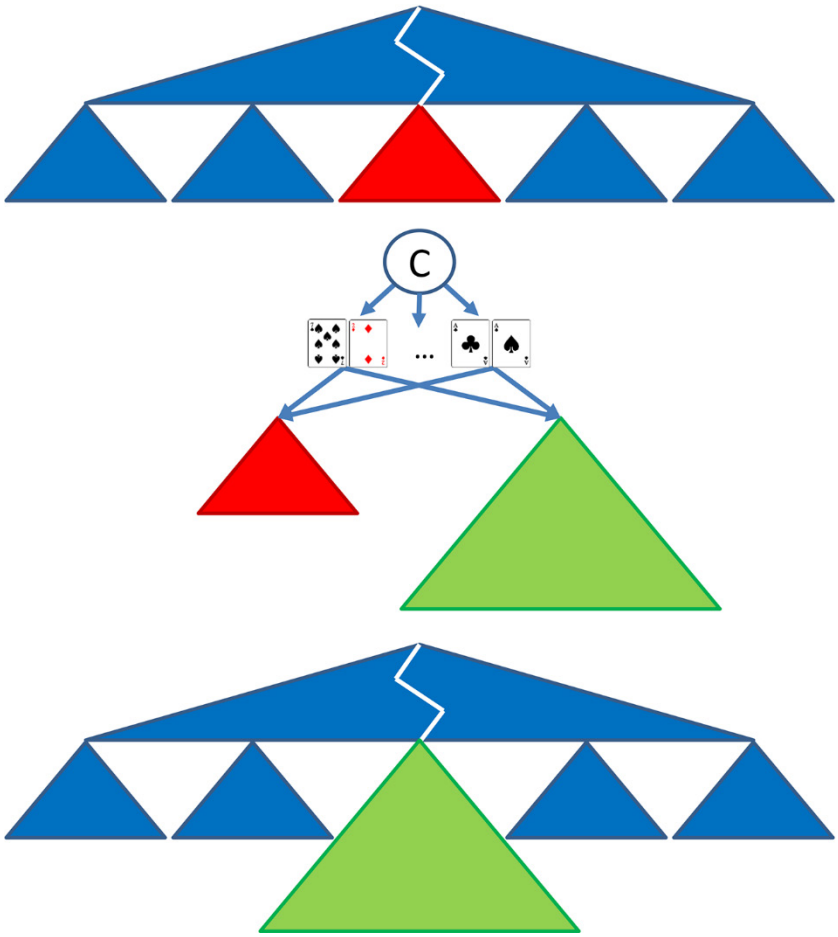


Fig. 1. Subgame solving. Top: A subgame is reached during play. Middle: A more detailed strategy for that subgame is determined by solving an augmented subgame, in which on each iteration the opponent is dealt a random poker hand and given the choice of taking the expected value of the old abstraction (red), or of playing in the new, finer-grained abstraction (green) where the strategy for both players can change. This forces Libratus to make the finer-grained strategy at least as good as in the original abstraction against every opponent poker hand. Bottom: The new strategy is substituted in place of the old one.

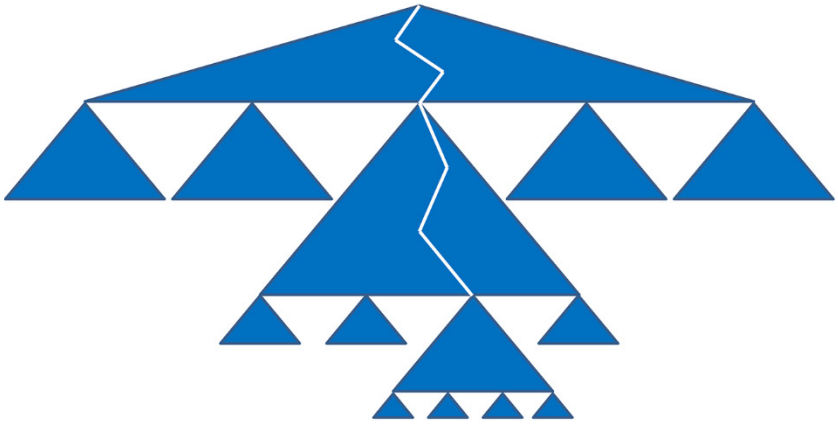


Fig. 2. A visualization of nested subgame solving. Every time a subgame is reached during play, a more detailed abstraction is constructed and solved just for that subgame, while fitting its solution within the overarching blueprint strategy.

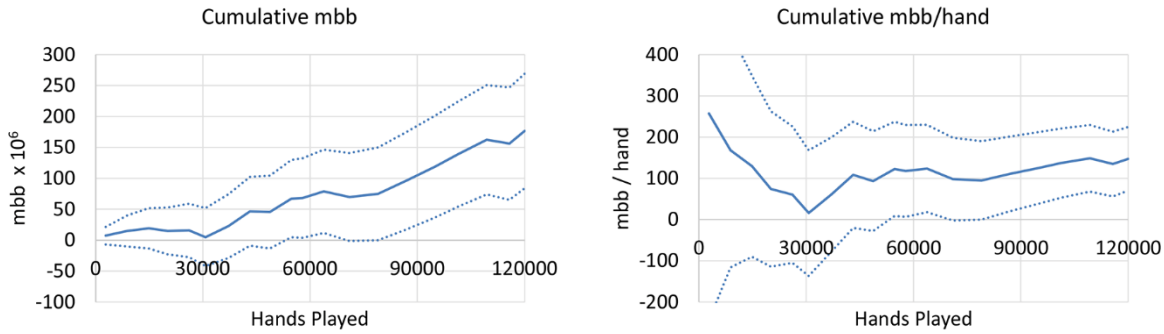


Fig. 3. Libratus performance against top humans. Shown are the results of the 2017 Brains vs. AI competition. The 95% confidence intervals (if the hands are treated as independent and identically distributed) are shown as dotted lines.

Table 1. Exploitability of subgame solving techniques on smaller poker variants.

Simplified Game:	Small 2-Round	Large 2-Round Hold'em	3-Round Hold'em
No subgame solving	91.3 mbb/hand	41.3 mbb/hand	346 mbb/hand
Unsafe subgame solving	5.51 mbb/hand	397 mbb/hand	79.3 mbb/hand
Safe subgame solving	22.6 mbb/hand	9.84 mbb/hand	72.6 mbb/hand

Table 2. Exploitability of nested subgame solving. Shown is the comparison to no nested subgame solving (which instead uses the leading action translation technique) in a small poker variant.

	Exploitability
No nested subgame solving	1,465 mbb/hand
Nested unsafe subgame solving	148 mbb/hand
Nested safe subgame solving	119 mbb/hand

Table 3. Head-to-head performance of Libratus. Shown are results for the Libratus blueprint strategy as well as forms of nested subgame solving against Baby Tartanian8 in HUNL.

	Performance against *Baby Tartanian8*
Blueprint	−8 ±15 mbb/hand
Blueprint with post-processing	18 ± 21 mbb/hand
On-tree nested subgame solving	59 ± 28 mbb/hand
Full nested subgame solving	63 ± 28 mbb/hand

HUMAN GENOMICS

The nature of nurture: Effects of parental genotypes

Augustine Kong,[1,2,3]* Gudmar Thorleifsson,[1] Michael L. Frigge,[1] Bjarni J. Vilhjalmsson,[4,5] Alexander I. Young,[1,2,6] Thorgeir E. Thorgeirsson,[1] Stefania Benonisdottir,[1] Asmundur Oddsson,[1] Bjarni V. Halldorsson,[1] Gisli Masson,[1] Daniel F. Gudbjartsson,[1,3] Agnar Helgason,[1,7] Gyda Bjornsdottir,[1] Unnur Thorsteinsdottir,[1,8] Kari Stefansson[1,8]*

Sequence variants in the parental genomes that are not transmitted to a child (the proband) are often ignored in genetic studies. Here we show that nontransmitted alleles can affect a child through their impacts on the parents and other relatives, a phenomenon we call "genetic nurture." Using results from a meta-analysis of educational attainment, we find that the polygenic score computed for the nontransmitted alleles of 21,637 probands with at least one parent genotyped has an estimated effect on the educational attainment of the proband that is 29.9% ($P = 1.6 \times 10^{-14}$) of that of the transmitted polygenic score. Genetic nurturing effects of this polygenic score extend to other traits. Paternal and maternal polygenic scores have similar effects on educational attainment, but mothers contribute more than fathers to nutrition- and heath-related traits.

How the human genome (nature) and the environment (nurture) work together to shape members of our species is a fundamental question, and any insights into this topic would be an important milestone. One challenge encountered by those who aspire to shed light on this matter is the lack of independence between the genome and the environment; thus, models that fail to account for this limitation are incomplete. Here we demonstrate how the genomes of close relatives—parents and siblings—can affect the proband through their contributions to the environment.

In animal studies, it is well established that alleles in a parent that are not transmitted to the offspring can nonetheless influence the offspring's phenotypes (*1*, *2*). Most examples involve effects manifested at the fetal stage, at which only the nontransmitted maternal alleles are relevant. In humans, the nontransmitted maternal alleles have been used to examine the potential causal relationships between the state of the mother during pregnancy and the outcomes of the child (*3*, *4*). Here, for humans, we consider an alternative causal path where both paternal and maternal nontransmitted alleles can have effects that are mostly manifested after birth. A sequence variant that affects the phenotype of an individual is also likely to affect the parent from whom it was inherited (Fig. 1A). For some phenotypes, the state of a parent can influence the state of its child. This gives rise to a situation in which a child's phenotype is influenced not only by the transmitted paternal and maternal alleles (T_P and T_M) (Fig. 1A) but also by the alleles that were not transmitted (NT_P and NT_M). A good example is educational attainment (EA) (*5*, *6*): The EA of parents provides an environmental effect for children, but one that has a genetic component (*7*, *8*). We call this phenomenon "genetic nurture." The transmitted and nontransmitted alleles (Fig. 1A) both exert effects on the parents, and thus both induce genetic nurturing effects. The effect of the transmitted allele includes both its direct effect on the proband and its effect manifested through nurturing from blood relatives. Because the amount of trait variance explained is proportional to the square of effect size, genetic nurture could have a larger impact on variance explained through the transmitted alleles (by magnifying the direct effect) than the nontransmitted alleles. However, data on the nontransmitted alleles are needed to separate the genetic nurturing effects from the direct effects of the transmitted alleles. Specifically, $\hat{\theta}_T$ (transmitted) and $\hat{\theta}_{NT}$ (nontransmitted) denote the respective estimated effects of the alleles when the paternal and maternal alleles are grouped together. Denoting the direct effect as δ, we propose to estimate it by $\hat{\delta} = (\hat{\theta}_T - \hat{\theta}_{NT})$. By calculating the difference, genetic nurturing effects and other potential confounding effects induced by population structure and assortative mating (*9*, *10*) (see below) are cancelled out. Even though the implementations are different, this approach is related to the transmission-disequilibrium test (TDT) (*11*, *12*), as both use nontransmitted alleles as controls (*13*). However, the potential effects of the nontransmitted alleles are ignored in the TDT. Mathematically, genetic nurture is a form of associative (or indirect) genetic effect, as defined by the animal-breeding literature (*2*). Genetic nurture is not limited to effects manifested through the phenotypes of the parents, as additional contributions (albeit probably substantially smaller

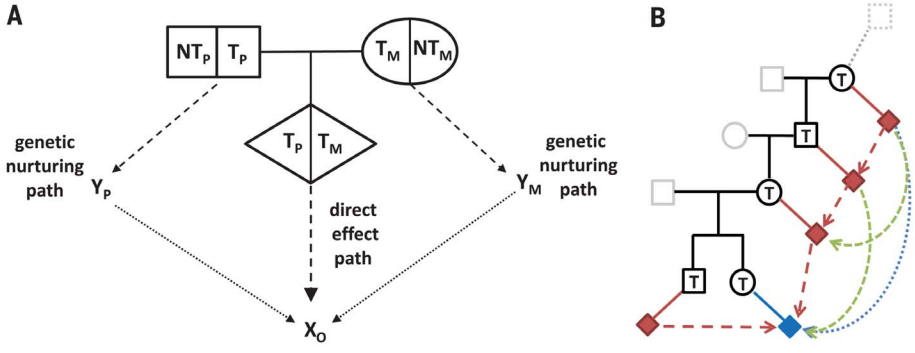


Fig. 1. Direct genetic effect and genetic nurturing effect. (**A**) Alleles at an autosomal site carried by a parents-offspring trio are labeled with respect to the offspring (proband). T_P and T_M denote, respectively, the alleles transmitted from the father and the mother to the proband, and NT_P and NT_M denote the paternal and maternal alleles that are not transmitted. The transmitted alleles can influence the phenotype of the offspring, X_O, through a direct path. The alleles of the parents, both transmitted and nontransmitted, can influence the parents' phenotypes, Y_P and Y_M, and through them may have a nurturing effect on X_O. This pathway combines a genetic effect (T_P, NT_P, T_M, and NT_M) on Y_P and Y_M with a nurturing effect (Y_P and Y_M) on X_O. Note that although X_O is often an individual trait of interest, Y would include a much broader set of phenotypes and is not completely known. (**B**) Red diamonds denote phenotypes of relatives; the blue diamond denotes the phenotype of the proband. Using the maternally transmitted allele as an example (denoted by T), we highlight that, in addition to the parents, the genetic nurturing effect can be manifested through the phenotypes of older ancestors and nonancestors such as siblings.

[1]deCODE genetics/Amgen, 101 Reykjavik, Iceland. [2]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford OX3 7LF, UK. [3]School of Engineering and Natural Sciences, University of Iceland, 101 Reykjavik, Iceland. [4]Bioinformatics Research Centre, Aarhus University, 8000 Aarhus, Denmark. [5]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA. [6]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. [7]Department of Anthropology, University of Iceland, 101 Reykjavik, Iceland. [8]Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland.
***Corresponding author. Email: augustine.kong@bdi.ox.ac.uk (A.K.); kstefans@decode.is (K.S.)**

Table 1. Decomposition of the observed effect of the polygenic score into direct, genetic nurturing, and confounding effects. Traits: educational attainment (EA), age at first child (AGFC), high-density lipoprotein level (HDL), body mass index (BMI), fasting glucose level (FG), height (HT), cigarettes per day for smokers (CPD), and composite health trait (HLTH). Traits are standardized to have a variance of 1. *N*: number of probands with at least one parent genotyped; N_{NTP}: number with father genotyped; N_{NTM}: number with mother genotyped. $\hat{\theta}_T$ and $\hat{\theta}_{NT}$: estimated effects of the polygenic scores computed for the transmitted and nontransmitted alleles, respectively, when they are analyzed jointly. $\hat{\delta} = (\hat{\theta}_T - \hat{\theta}_{NT})$: estimated direct effect of the polygenic score. R^2: estimated variance accounted for by the transmitted polygenic score, which captures both the direct effect and the genetic nurturing effect. R^2_δ: estimated variance accounted for by the direct effect alone. These fractions of variance explained are for trait values adjusted for sex, yob (year of birth), and PCs. Corresponding values for unadjusted trait values would be somewhat smaller (*13*). $\hat{\phi}_\delta$, $\hat{\eta}$, and $\hat{\phi}_\eta$: estimates, respectively, of the assortative mating–induced confounding effect for the direct effect component, the genetic nurturing effect, and the confounding effect of the genetic nurturing component.

Trait	*N*	N_{NTP}	N_{NTM}	Transmitted T ($T = T_P + T_M$)			Nontransmitted NT ($NT = NT_P + NT_M$)		R^2_δ (%)	$\hat{\delta}/\hat{\theta}_T$	$\hat{\phi}_\delta/\hat{\theta}_T$	$\hat{\eta}/\hat{\theta}_T$	$\hat{\phi}_\eta/\hat{\theta}_T$
				$\hat{\theta}_T$	*P*	R^2 (%)	$\hat{\theta}_{NT}$	*P*					
EA	21637	13948	19012	0.223	1.6×10^{-174}	4.98	0.067	1.6×10^{-14}	2.45	0.701	0.046	0.224	0.029
AGFC	54372	35294	47052	0.108	9.7×10^{-110}	1.17	0.039	2.9×10^{-13}	0.48	0.640	0.052	0.264	0.043
HDL	46872	30855	40788	0.065	9.0×10^{-29}	0.42	0.027	6.0×10^{-6}	0.14	0.586	0.046	0.319	0.050
BMI	39078	26433	34533	−0.060	1.0×10^{-22}	0.36	−0.017	0.0077	0.19	0.718	0.055	0.197	0.030
FG	34767	22959	30222	−0.051	7.6×10^{-18}	0.26	−0.018	0.0059	0.11	0.655	0.052	0.252	0.040
HT	39270	26563	34703	0.052	6.6×10^{-14}	0.28	0.030	1.5×10^{-5}	0.05	0.422	0.031	0.476	0.071
CPD	18887	12371	16589	−0.055	1.4×10^{-12}	0.31	−0.030	5.3×10^{-4}	0.06	0.461	0.035	0.439	0.066
HLTH	62328	41996	54546	0.082	2.7×10^{-60}	0.67	0.033	8.9×10^{-11}	0.23	0.592	0.051	0.305	0.052

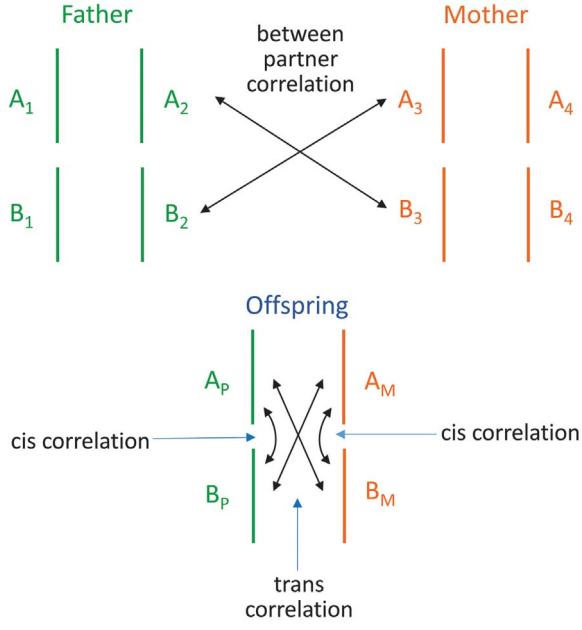


Fig. 2. Correlation and confounding induced by assortative mating. An example of two loci, A and B, contributing to the phenotype. Through assortative mating, alleles in the father become correlated with alleles in the mother. Consequently, the transmitted paternal alleles (A_P and B_P) become correlated with the maternally transmitted alleles (A_M and B_M). This correlation between alleles with different parental origins is referred to as trans correlation, whereas correlation between alleles with the same parental origins (e.g., A_P and B_P) is referred to as cis correlation. When A_P/A_M and B_M/B_P are correlated, association analysis between the phenotype and A alone will also capture part of the effect of B.

ones) may go through grandparents and great-grandparents, for example (Fig. 1B). This study takes advantage of our human data to empirically examine the magnitudes of such effects for traits such as EA.

Estimating direct effects

To maximize the power to detect the effects of the nontransmitted alleles, we used 618,762 single-nucleotide polymorphisms (SNPs) spanning the genome to construct polygenic scores (*14*). The per-locus allele-specific weightings for the polygenic scores were derived from applying LDpred (*15*) to the results of a large genome-wide association study (*8*) (GWAS) of EA measured in years of education, with Icelandic data removed (*13*). The first analysis focused on 21,637 Icelandic probands, born between 1940 and 1983 (9139 males, 12,498 females), with EA data and at least one parent genotyped (Table 1). Because we could establish the parent of origin of the transmitted alleles (*16*), the nontransmitted allele from a genotyped parent was easily determined. $poly_{TP}$ and $poly_{TM}$ represent the polygenic scores computed from the transmitted paternal and maternal alleles, respectively, and $poly_{NTP}$ and $poly_{NTM}$ denote the corresponding polygenic scores for the nontransmitted alleles. To maximize power, we start by providing the results for $poly_T = poly_{TP} + poly_{TM}$ and $poly_{NT} = poly_{NTP} + poly_{NTM}$. Here, $poly_{TP}$ and $poly_{TM}$ are scaled so that $poly_T$ has a mean of 0 and a variance of 1, and the trait EA is standardized to have a variance of 1. $poly_{NTP}$ and $poly_{NTM}$ were similarly calculated, and a 0 was imputed when the parent was not genotyped (*13*). Associations between EA and the polygenic scores computed from a joint analysis of $poly_T$ and $poly_{NT}$ that adjusts for sex, year of birth (yob) up to the cubic term, interactions between sex and yob, and 100 principal components (PCs) (*13*) are presented in Table 1. The estimated effect of $poly_T$, $\hat{\theta}_T$, is 0.223 and significant [$P = 1.6 \times 10^{-174}$, calculated with genomic control adjustment (*13*, *17*)]. Because both $poly_T$ and EA are standardized, the estimated fraction of the trait variance explained by $poly_T$ is $\hat{\theta}_T^2 = (0.223)^2 = 4.98\%$ (R^2 in Table 1). However, the estimated effect of $poly_{NT}$, $\hat{\theta}_{NT} = 0.067$, is also significant ($P = 1.6 \times 10^{-14}$). Thus, the estimated direct effect of $poly_T$, $\hat{\delta} = (\hat{\theta}_T - \hat{\theta}_{NT}) = 0.157$, explains only $R^2_\delta = 0.157^2 = 2.45\%$ of the trait variance, approximately one-half of R^2. Noting that $R^2_\delta/R^2 = (\hat{\delta}/\hat{\theta}_T)^2$, the value of $\hat{\delta}/\hat{\theta}_T$ is determined (Table 1). In addition to the polygenic scores, individual results for 120 genome-wide significant SNPs ($P < 5 \times 10^{-8}$) in the Iceland-excluded meta-analysis are provided (table S1). Fifteen of the 120 SNPs (12.5%) have a one-tailed *P* value that is <0.05 for the nontransmitted alleles, which is more than that expected from noise [$P = 1.5 \times 10^{-3}$ (*13*)]. The average estimated effect of the nontransmitted alleles is 34.2% of that of the transmitted alleles. These results are consistent with previous observations that within-family EA effects calculated for dizygotic twins tended to be

Table 2. Parent-of-origin–specific effects of the polygenic scores. Traits are as defined in Table 1. $\hat{\theta}_{TP}$, $\hat{\theta}_{TM}$, $\hat{\theta}_{NTP}$, and $\hat{\theta}_{NTM}$: estimates of the effect of the polygenic scores $poly_{TP}$, $poly_{TM}$, $poly_{NTP}$, and $poly_{NTM}$, respectively. $\hat{\eta}_P$ and $\hat{\eta}_M$: estimates of the paternal and maternal genetic nurturing effects, respectively.

Trait	Transmitted				Nontransmitted				Estimate of $\eta_M - \eta_P$	P	$\hat{\eta}_M/\hat{\eta}_P$
	$\hat{\theta}_{TP}$	P	$\hat{\theta}_{TM}$	P	$\hat{\theta}_{NTP}$	P	$\hat{\theta}_{NTM}$	P			
EA	0.214	1.0×10^{-89}	0.232	9.9×10^{-103}	0.066	5.2×10^{-7}	0.067	3.6×10^{-9}	0.011	0.31	1.26
AGFC	0.100	1.4×10^{-52}	0.116	1.9×10^{-68}	0.034	2.7×10^{-5}	0.043	1.2×10^{-9}	0.013	0.067	1.59
HDL	0.062	5.0×10^{-16}	0.068	7.8×10^{-19}	0.013	0.13	0.037	2.0×10^{-6}	0.014	0.077	2.05
BMI	−0.059	5.2×10^{-13}	−0.062	1.2×10^{-13}	−0.019	0.055	−0.016	0.062	−0.000	0.98	1.02
FG	−0.043	1.0×10^{-7}	−0.059	4.1×10^{-13}	−0.011	0.27	−0.023	0.0073	−0.014	0.090	3.99
HT	0.035	8.8×10^{-5}	0.070	1.3×10^{-14}	0.027	0.0082	0.033	3.5×10^{-4}	0.023	0.011	2.85
CPD	−0.042	1.3×10^{-4}	−0.069	2.6×10^{-10}	−0.036	0.0071	−0.025	0.028	−0.011	0.33	1.63
HLTH	0.070	1.4×10^{-26}	0.093	1.0×10^{-44}	0.026	7.5×10^{-4}	0.039	8.4×10^{-9}	0.019	0.0048	2.32

smaller than the standard GWAS effect estimates (*7*, *8*).

Assortative mating and estimating the genetic nurturing effect

We designate η to denote the magnitude of the genetic nurturing effect. Even though our analyses have adjustment for 100 PCs, which should have eliminated much of the population stratification–induced confounding, $\hat{\theta}_{NT}$ can still capture effects other than η. When there is assortative mating with respect to the genetic component underlying EA (*10*), a subtle confounding effect may result. Figure 2 illustrates a simple scenario in which the phenotype is assumed to be influenced by two loci: A and B. If there is assortative mating in the parents' generation, it would lead to correlation of alleles between partners; for instance, the A alleles of the father (A_1 and A_2 in Fig. 2) will be correlated with the B alleles of the mother (B_3 and B_4) and vice versa. Consequently, the paternally transmitted A allele A_P will be positively correlated with the maternally transmitted B allele B_M, and A_M will be correlated with B_P. This correlation between alleles inherited from different parents is referred to as trans correlation, whereas the correlation between alleles inherited from the same parent (e.g., A_P and B_P) is referred to as cis correlation. This assortative mating–induced correlation differs from correlation between markers that are close physically, that is, within the same linkage-disequilibrium block. The latter correlation is mainly driven by the cis component, whereas the assortative mating–induced correlation could be dominated by the trans component. If trait association is calculated for locus A individually, the observed effect will capture both the effect of locus A and part of the effect of locus B. We let ϕ_δ denote this added confounding effect. Similarly, assortative mating would also lead the A alleles to capture some of the nurturing effect of locus B, an effect denoted by ϕ_η. Under our model assumptions (*13*)

$$\phi_\eta/\eta = 2 \times (\phi_\delta/\delta)$$

The factor of 2 arises because the nontransmitted alleles have the same nurturing effects as the transmitted ones, and thus the transmitted and nontransmitted A alleles are capturing, through correlation, the nurturing effects of both the transmitted and nontransmitted B alleles. Additionally, we have the decompositions

$$E[\hat{\theta}_T] = \delta + \phi_\delta + \eta + \phi_\eta$$

and

$$E[\hat{\theta}_{NT}] = \phi_\delta + \eta + \phi_\eta$$

where $E[\ ]$ denotes expectation. Because both the transmitted and nontransmitted A alleles capture the confounding effects, $\hat{\delta} = (\hat{\theta}_T - \hat{\theta}_{NT})$ remains an appropriate estimate of the direct effect δ. Locus A and locus B in Fig. 2 can be generalized to represent two nonoverlapping sets of loci. For our study, we think of locus A as the EA polygenic score, whereas locus B represents the genetic component of EA that is statistically orthogonal to locus A (under a scenario of no assortative mating). The mathematical relationships highlighted above continue to hold for the polygenic scores, either exactly or approximately. By using a method for estimating heritability that also incorporates data on the nontransmitted alleles (*18*), we estimate the full genetic component of EA to have a direct effect that explains 17.0% of the variance of EA. In other words, $poly_T$ is estimated to be 2.45/17.0 = 14.4% of the full genetic component, whereas the remaining 85.6% corresponds to the B components. From this estimate, we extrapolate the correlations observed between the paternal polygenic scores ($poly_{TP}$ and $poly_{NTP}$) and the maternal polygenic scores ($poly_{TM}$ and $poly_{NTM}$) to estimate the correlations between them and the unobserved B components (*13*). From the latter, ϕ_δ/δ and ϕ_η/η are estimated as 0.065 and 0.130, respectively. For this calculation, we avoided making the assumption that assortative mating between parents was manifested only through correlation of their EAs, which would have led to lower estimates for the ϕ values (*13*). From these estimates and the above equations, $\hat{\phi}_\delta$, $\hat{\eta}$, and $\hat{\phi}_\eta$ were computed and presented in Table 1 as fractions of $\hat{\theta}_T$. For EA, $\hat{\eta}$ accounts for ~75% of the value of $\hat{\theta}_{NT}$ and $\hat{\eta}$ is 31.9% of $\hat{\delta}$. Finally, we note that assortative mating occurring before the parents' generation could lead to additional confounding. However, this effect appears to be negligible in our study, as after adjustment for 100 PCs, the within-parent correlation of the transmitted and nontransmitted polygenic scores is actually negative (but $P > 0.05$) (*13*).

Direct and nurturing effects on other traits

The EA polygenic score is associated with other quantitative traits in our database. Among them, those with the strongest statistical significance (Table 1) are age at first child (AGFC) (*19*), high-density lipoprotein level (HDL) (*20*), body mass index (BMI) (*21*), fasting glucose level (FG) (*22*), height (HT) (*23*), and cigarettes smoked per day by smokers (CPD) (*24*). The effects of the transmitted and nontransmitted EA polygenic scores on these phenotypes were estimated as before for the EA phenotype (Table 1). Although the fraction of variance explained by $poly_T$ (R^2) is smaller than that for EA, the effect of $poly_{NT}$ is statistically significant. Moreover, except for BMI, the ratio $\hat{\eta}/\hat{\delta}$ is higher for these traits than for EA and exceeds 1 for HT.

Parent of origin

Table 2 provides the estimated effects of $poly_{TP}$, $poly_{TM}$, $poly_{NTP}$, and $poly_{NTM}$ separately (*13*). For EA, $\hat{\theta}_{NTP}$, the estimated effect of $poly_{NTP}$, is significant ($P = 5.2 \times 10^{-7}$), and its value is nearly identical to that of $\hat{\theta}_{NTM}$ (the higher P value for $poly_{NTP}$ is due to fewer fathers genotyped than mothers). This indicates that the effect observed for $poly_{NT}$ is not driven by epigenetic effects such as imprinting or genetic interactions between fetus and mother in the womb and does capture a genetic nurturing effect [also see tables S2 and S3, which have results for polygenic scores calculated without SNPs in imprinted regions (*25*)]. However, even with both parents contributing to genetic nurture, the magnitude of the effect can differ between fathers and mothers. We designate η_P and η_M to denote the paternal and maternal genetic nurturing effects, respectively. Because the transmitted alleles also contribute to the

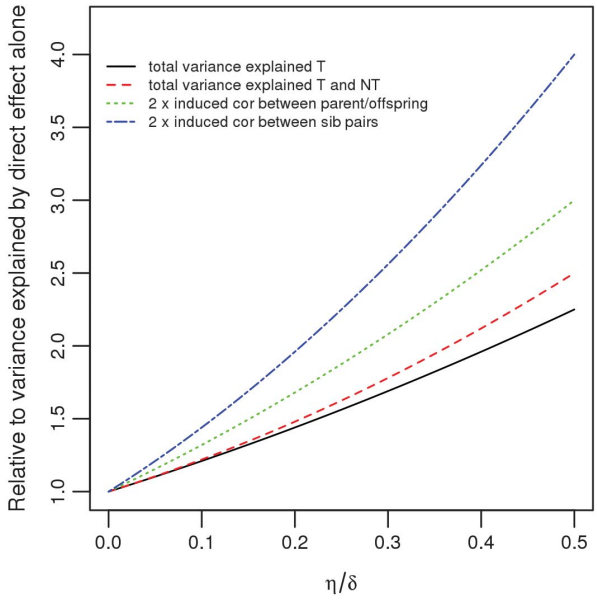


Fig. 3. Variance explained and induced correlations between phenotypes of parent-offspring and siblings. Results are displayed as a function of the ratio η/δ. The *y* axis is the relative amplification; that is, various measures relative to what can be accounted for by the direct effect alone, the latter proportional to δ^2. The total variance explained by the transmitted alleles is proportional to $(\delta + \eta)^2$ [the plotted curve is hence $(\delta + \eta)^2/\delta^2$], whereas the total variance explained by the transmitted alleles plus the nontransmitted alleles is proportional to $(\delta + \eta)^2 + \eta^2$. Formulas for the induced parent-offspring and sibling correlations are derived (*13*). η, magnitude of genetic nurturing effect; δ, direct effect; T, transmitted; NT, nontransmitted.

nurturing effect, we use a weighted average of $(\hat{\theta}_{TM} - \hat{\theta}_{TP})$ and $(\hat{\theta}_{NTM} - \hat{\theta}_{NTP})$, with weights proportional to the inverse square of the standard error (*13*), to estimate $(\eta_M - \eta_P)$ (Table 2). Combining this estimate with $\hat{\eta}$ from Table 1, considered as an estimate of a weighted average of η_P and η_M with weights proportional to the numbers of fathers and mothers genotyped, we calculated individual estimates of η_P and η_M (*13*), denoted by $\hat{\eta}_P$ and $\hat{\eta}_M$, and the ratio $\hat{\eta}_M/\hat{\eta}_P$ (Table 2). For EA, $(\eta_M - \eta_P)$ is estimated to be 0.011, but it is not significantly different from zero ($P = 0.31$)—that is, the ratio $\hat{\eta}_M/\hat{\eta}_P = 1.26$ is not significantly different from 1. For all of the other six traits, $\hat{\eta}_M/\hat{\eta}_P > 1$ but was significant only for HT ($\hat{\eta}_M/\hat{\eta}_P = 2.85$, $P = 1.1 \times 10^{-2}$). HDL and FG have *P* values between 0.05 and 0.10. To increase power, for individuals for whom we had data for one or more of the five health- and nutrition-related traits (HDL, BMI, FG, HT, and CPD), a composite health trait (HLTH) was constructed by taking the sum of the standardized values of the available traits (positive signs for HDL and HT; negative signs for BMI, FG, and CPD) and dividing it by the square root of the number of trait values summed. It was then standardized to have a variance of 1. For HLTH, $\hat{\theta}_{NT}$ has a larger value than that for the individual health- and nutrition-related traits and is highly significant ($P = 8.9 \times 10^{-11}$) (Table 1). Both $\hat{\theta}_{NTP}$ and $\hat{\theta}_{NTM}$ are significant, but $\hat{\eta}_M/\hat{\eta}_P = 2.32$ with a *P* value of 4.8×10^{-3} (Table 2). This supports the notion that mothers have a stronger nurturing effect than fathers on the health of the child.

Variance explained and effects of siblings

The existence of genetic nurture complicates the estimation and interpretation of heritability (*18*). For example, maternal effects have been shown to affect heritability estimates from animal-breeding data (*26*). Though distinct from the direct effect of inherited genetic variants, we demonstrate here how genetic nurture can be measured and taken into consideration. If there are two uncorrelated variants of the same frequency, one having a direct effect δ only and the other having a nurturing effect η only, then the variance explained is proportional to $\delta^2 + \eta^2$. By comparison, if one variant has both effects, then the variance explained is proportional to $(\delta + \eta)^2 = \delta^2 + 2\delta\eta + \eta^2$ (Fig. 3), with the extra $2\delta\eta$ term. Moreover, $(\delta + \eta)^2$ captures the effect of the transmitted allele(s) only; the phenotypic variance accounted for by the transmitted and nontransmitted alleles together is proportional to $(\delta + \eta)^2 + \eta^2$ (Fig. 3). With EA, $\hat{\eta}/\hat{\delta} = 0.319$, $(\hat{\delta} + \hat{\eta})^2/\hat{\delta}^2 = 1.74$, and $[(\hat{\delta} + \hat{\eta})^2 + \hat{\eta}^2]/\hat{\delta}^2 = 1.84$. Assuming that the direct effect alone explains 17.0% of the variance, the variance explained by the transmitted alleles with the nurturing effects included increases to 17.0% × 1.74 = 29.6%. Additionally including the nontransmitted alleles would increase the variance explained to 17.0% × 1.84 = 31.3%. The genetic nurturing effect not only magnifies the variance explained, it also induces an even larger amplification of the phenotypic correlations of parents and offspring and of siblings (Fig. 3) (*13*). Also worth noting is that the $2\delta\eta$ term highlighted above does not exist for adopted children, as then both alleles of a parent would be nontransmitted.

Genetic nurture could go through a sibling (Fig. 1B) if, as proposed (*27*), the phenotypes of the proband are influenced by the phenotypes or behavior of a sibling. On the basis of the genealogy, for each EA proband who has at least one sibling, the sibling most likely to have the biggest effect on the proband was identified as follows. If the proband has older siblings, the older sibling with a yob closest to the proband was selected (monozygotic twins were excluded, but we count a dizygotic twin of the proband as an older sibling). If the proband is the eldest child, a younger sibling with the closest yob was chosen. There are 7798 probands whose chosen sibling is genotyped and whose parents are both genotyped. A polygenic score, denoted by $poly_{TS}$, was computed using the alleles transmitted from the parents to the sibling. The EA of the proband was then regressed on $poly_T$, $poly_{NT}$, and $poly_{TS}$ jointly. The effect of $poly_{TS}$ is significant ($P = 0.015$) and is estimated to be 24.1% (95% confidence interval: 4.7 to 43.6%) of the direct effect. The uncertainty is large because $poly_{TS}$ is strongly correlated with $poly_T$ and $poly_{NT}$. One compensation is that, having adjusted for both $poly_T$ and $poly_{NT}$, the estimated effect of $poly_{TS}$ is free of confounding from assortative mating.

Heritability is defined as the fraction of phenotypic variance explained by direct effects alone. The presence of parental genetic nurture introduces bias to estimates of heritability from GREML (genomic relatedness–based restricted maximum likelihood)–type methods (*28*), such as those embodied in the software package GCTA (*29*), that use correlations due to transmitted alleles without distinction between direct genetic effects and genetic nurturing effects (*18*). By contrast, heritability estimates from comparing correlations between monozygotic versus dizygotic twins (*30*) are unaffected as the effects of parental genetic nurture are cancelled out. However, when genetic nurturing effects that go through the phenotypes of a sibling or twin are present, both twin-based heritability estimates (*31*) and estimates from GREML-type methods are affected.

The nature of genetic nurture and other polygenic scores

To further use the EA trait data, we performed analyses that treated the nontransmitted polygenic score of a genotyped parent as missing if the EA of that parent was unknown. For these data, (unadjusted) estimates of θ_{NT} were calculated as before (table S4). Also given are estimates of θ_{NT} adjusted for the EAs of the parents, obtained by adding the latter to the explanatory variables in the regressions. For EA, AGFC, HT, and HLTH, the adjusted estimate remains significant ($P < 0.005$), and the ratio of the adjusted versus unadjusted estimate is, respectively, 47.6, 63.0, 80.3, and 68.6%. This indicates that the EA of the parent is an important part of the parental phenotypes (Y in Fig. 1A) through which genetic nurture operates, but it is far from all of it. The EA polygenic score is likely associated with intelligence, conscientiousness, and future planning. Parents with a high score enhance the nurturing of their offspring through many behaviors, not exclusively through their own EA.

To contrast the results presented for the polygenic score constructed from a GWAS of EA (EA polygenic score), we examined polygenic scores constructed from GWASs of HT (*32*) (HT polygenic score) and BMI (*33*) (BMI polygenic score).

(Results corresponding to Table 1 are in tables S5 and S6.) Noting that the HT and BMI polygenic scores are, respectively, positively (r = 0.087) and negatively correlated (r = −0.146) with the EA polygenic score, we computed HT and BMI polygenic scores adjusted for the EA polygenic score by regressing the former on the latter and calculating the residuals (tables S7 and S8). Whereas the unadjusted nontransmitted polygenic score has a few significant associations (tables S5 and S6), with adjustment (tables S7 and S8) the only significant effect of the nontransmitted polygenic score is between the HT trait and the nontransmitted HT polygenic score. Furthermore, most of this observed effect is estimated to be due to confounding from assortative mating.

Discussion

Through the study of the nontransmitted alleles, we demonstrated that genetic nurturing effects exist and can have an impact on variance explained. These results also reveal that the observed effects from GWAS do not necessarily reflect direct effects alone. They can be amplified by genetic nurturing effects and, to a lesser extent, assortative mating–induced confounding. Owing to power considerations, we mostly studied variants as an aggregate. However, given the complexity of the EA trait (*6*) and our observed effects of the EA polygenic score on other traits, for individual variants, the ratio of the genetic nurturing effect versus the direct effect must have variations both between and within traits. Thus, we should aim to gather enough data to perform GWAS with the nontransmitted alleles. This would add insight into the pathway(s) through which the effect of an individual variant is manifested, as well as enable a better understanding of some pleiotropic effects (*34*).

Although genes have been shown to affect the environment (*24*, *35*, *36*), the contribution of a genetic effect manifested through nurturing has mostly been ignored in GWAS. Results here highlight the importance of family data.

Our focus has been on genetic nurture in one direction, but the effects are likely to be bidirectional. For a parent-offspring pair, the magnitude of the effect in the direction of parent to offspring is likely to dominate the effect in the opposite direction. However, with siblings and twins, the effects would be reciprocal.

Our analyses implicitly assume that direct genetic effects and genetic nurturing effects are additive, but interactive effects could certainly exist and further complicate the interpretation of observed effects. Moreover, alleles other than those in the parents can also have an effect; for example, the genetic makeup of the population of the probands could also be an important environmental contributor to their phenotypes.

REFERENCES AND NOTES

1. B. Griffing, *Aust. J. Biol. Sci.* **20**, 127–139 (1967).
2. M. Lynch, B. Walsh, *Genetics and Analysis of Quantitative Traits* (Sinauer Associates, 1998).
3. G. Zhang *et al*., *PLOS Med.* **12**, e1001865 (2015).
4. R. C. Richmond *et al*., *PLOS Med.* **14**, e1002221 (2017).
5. N. E. Hill, D. F. Tyson, *Dev. Psychol.* **45**, 740–763 (2009).
6. E. Krapohl *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15273–15278 (2014).
7. C. A. Rietveld *et al*., *Science* **340**, 1467–1471 (2013).
8. A. Okbay *et al*., *Nature* **533**, 539–542 (2016).
9. D. S. Falconer, T. F. C. Mackay, *Introduction to Quantitative Genetics* (Pearson, 1996).
10. M. R. Robinson *et al*., *Nat. Hum. Behav.* **1**, 0016 (2017).
11. R. S. Spielman, R. E. McGinnis, W. J. Ewens, *Am. J. Hum. Genet.* **52**, 506–516 (1993).
12. D. B. Allison, *Am. J. Hum. Genet.* **60**, 676–690 (1997).
13. Materials and methods are available as supplementary materials.
14. N. R. Wray, M. E. Goddard, P. M. Visscher, *Genome Res.* **17**, 1520–1528 (2007).
15. B. J. Vilhjálmsson *et al*., *Am. J. Hum. Genet.* **97**, 576–592 (2015).
16. A. Kong *et al*., *Nature* **462**, 868–874 (2009).
17. B. K. Bulik-Sullivan *et al*., *Nat. Genet.* **47**, 291–295 (2015).
18. A. Young *et al*., Estimating heritability without environmental bias. bioRxiv 218883 [Preprint] 14 November 2017.
19. N. Barban *et al*., *Nat. Genet.* **48**, 1462–1472 (2016).
20. A. Helgadottir *et al*., *Nat. Genet.* **48**, 634–639 (2016).
21. G. Thorleifsson *et al*., *Nat. Genet.* **41**, 18–24 (2009).
22. J. Flannick *et al*., *Nat. Genet.* **46**, 357–363 (2014).
23. D. F. Gudbjartsson *et al*., *Nat. Genet.* **40**, 609–615 (2008).
24. T. E. Thorgeirsson *et al*., *Nature* **452**, 638–642 (2008).
25. S. Benonisdottir *et al*., *Nat. Commun.* **7**, 13490 (2016).
26. K. Meyer, *Livest. Prod. Sci.* **31**, 179–204 (1992).
27. C. Nicoletti, B. Rabe, "Sibling spillover effects in school achievement" [Discussion Paper no. 8615, Institute for the Study of Labor (IZA), 2014].
28. J. Yang *et al*., *Nat. Genet.* **42**, 565–569 (2010).
29. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, *Am. J. Hum. Genet.* **88**, 76–82 (2011).
30. T. J. Polderman *et al*., *Nat. Genet.* **47**, 702–709 (2015).
31. M. J. Rietveld, D. Posthuma, C. V. Dolan, D. I. Boomsma, *Behav. Genet.* **33**, 247–255 (2003).
32. A. R. Wood *et al*., *Nat. Genet.* **46**, 1173–1186 (2014).
33. A. E. Locke *et al*., *Nature* **518**, 197–206 (2015).
34. B. Bulik-Sullivan *et al*., *Nat. Genet.* **47**, 1236–1241 (2015).
35. R. Plomin, *Behav. Genet.* **44**, 629–638 (2014).
36. R. Dawkins, *The Extended Phenotype* (Oxford Univ. Press, 1982).

ACKNOWLEDGMENTS

We thank A. Okbay for providing the EA meta-analysis results with Icelandic data removed; J. Hirshhorn for pointing out that the nontransmitted alleles could be capturing some confounding effects due to assortative mating; and the GIANT consortium, A. Wood, and A. Locke for assisting us to obtain meta-analysis results for HT and BMI with Icelandic data removed. B.V.H. is an associate professor in the School of Science and Engineering at Reykjavík University. A summary of the data used in this manuscript is in table S9. Icelandic law allows for unimpeded sharing of summary-level data. However, the law does not allow the sharing of individual-level data on genotypes and phenotypes outside of Iceland.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6374/424/suppl/DC1
Materials and Methods
Tables S1 to S9
References (*37*–*43*)

22 May 2017; accepted 13 December 2017
10.1126/science.aan6877

ORGANIC CHEMISTRY

A platform for automated nanomole-scale reaction screening and micromole-scale synthesis in flow

Damith Perera,[1]* Joseph W. Tucker,[2] Shalini Brahmbhatt,[1] Christopher J. Helal,[2] Ashley Chong,[1] William Farrell,[1] Paul Richardson,[1]* Neal W. Sach[1]*

The scarcity of complex intermediates in pharmaceutical research motivates the pursuit of reaction optimization protocols on submilligram scales. We report here the development of an automated flow-based synthesis platform, designed from commercially available components, that integrates both rapid nanomole-scale reaction screening and micromole-scale synthesis into a single modular unit. This system was validated by exploring a diverse range of reaction variables in a Suzuki-Miyaura coupling on nanomole scale at elevated temperatures, generating liquid chromatography–mass spectrometry data points for 5760 reactions at a rate of >1500 reactions per 24 hours. Through multiple injections of the same segment, the system directly produced micromole quantities of desired material. The optimal conditions were also replicated in traditional flow and batch mode at 50- to 200-milligram scale to provide good to excellent yields.

In drug discovery programs, it is critical to rapidly synthesize project compounds with the potential to become new therapies, as well as to minimize time spent on ultimately nonoptimal analogs. As such, the application of new synthetic chemistry technologies can play a central role in the accelerated discovery of pharmaceutical agents. High-throughput experimentation (HTE) and flow chemistry (*1*) are enabling technologies (*2*) that often sit at opposite ends of the synthetic scale, with the former encompassing hundreds of micromole-scale batch-type reactions for optimization of conditions and the latter facilitating efficient bulk production of single compounds under wide temperature and pressure ranges (*3*). Although each technique has been advocated as a tool to expedite the drug discovery process, questions remain as to whether this has actually occurred, particularly as the costs involved with this endeavor continue to escalate (*4*). In the earliest phase of a medicinal chemistry program, materials are often a limiting factor, and to increase the coverage of chemical space within a screening campaign, it is necessary to decrease the scale of experimentation. Advanced liquid handling technologies and exclusion of air and moisture in glovebox environments have enabled robust reaction screens on >1-mg quantities of material at 50- to 100-µl volume per reaction. It has therefore become routine to run screens of hundreds of reactions that, in combination with high-throughput analytics such as ultra-performance liquid chromatography–mass spectrometry (UPLC-MS), achieve a rapid turnaround from design all the way through to results in 2 to 3 days, drastically altering the cost/benefit analysis of prospective reaction screening. This approach expands the potential structural diversity of investigated compounds and saves time and resource investment downstream in the development phase if efficient chemistries discovered through HTE are already in place (*5*).

In considering how to further leverage these methods, we were particularly inspired by Dreher, Cernak, and co-workers at Merck, who in a seminal publication provided an elegant solution to enable chemistry at nanomole scale using equipment and technology from biological-assay screening (*6*). Iterative reaction screening in 20-nl volumes successfully optimized a Pd-mediated Buchwald-Hartwig coupling to yield druglike fragments: 1536 reactions were evaluated in 2.5 hours with as little as 0.02 mg per reaction. However, despite this impressive work, there were still several limitations of this plate-based approach such as the need for nonvolatile solvents (e.g., dimethyl sulfoxide), the absence of heating to avoid solvent evaporation, and application of low-resolution mass spectrometry for analysis. We endeavored to develop a synthetic platform that would overcome these limitations, merging the best attributes of these approaches to screen large numbers of reactions at nanomole scale in a flow system under diverse solvent, temperature, and pressure conditions with the subsequent ability to synthesize hundreds of micromoles for possible biological testing in a highly automated fashion.

[1]Pfizer Worldwide Research and Development, La Jolla Laboratories, 10770 Science Center Drive, San Diego, CA 92121, USA. [2]Pfizer Worldwide Research and Development, Eastern Point Road, Groton, CT 06340, USA.
*Corresponding author. Email: sanjeewadamith.perera@pfizer.com (D.P.); paul.f.richardson@pfizer.com (P.R.); neal.sach@pfizer.com (N.W.S.)

The key objectives were to develop a fully automated system for HTE screening with flow chemistry technology that would (i) integrate inline high-resolution LC-MS analysis for real-time reaction monitoring; (ii) use diverse volatile and nonvolatile solvents; (iii) use ~0.05 mg of substrate per reaction to enable broad parameter space exploration with minimal material consumption; (iv) enable the preparation and analysis of up to 1500 reaction segments in a 24-hour period; (v) establish the capacity of the platform to directly scale up preferred conditions via multiple injections to produce 10- to 100-mg quantities of a specific compound; and (vi) show translation of nano-HTE conditions to both larger-scale batch and flow synthesis.

Solvent variation in flow HTE

Despite the advances in flow chemistry technology, its implementation for high-throughput reaction screening with multiple discrete (e.g., catalyst, ligand, base) and continuous (e.g., temperature, residence time, pressure) variables has been limited to date (*7*, *8*). Furthermore, all flow reaction screening systems have required the preparation of reagent stock solutions when solvent, a key reaction variable, is varied (*9*). This problem is a major hurdle to applying flow reaction screening to medicinal chemistry, where limited material does not support preparation of multiple stock solutions. As an alternative, we envisioned preparing concentrated reactant and reagent solutions in a suitable solvent for each reaction component. These would be injected into a carrier solvent with the intervals between injections being carefully monitored to prevent cross-contamination between the individual reaction segments (*10*). Diffusion of the injected segment (5 µl) into the carrier solvent (500 µl, 1:100) would result in sufficient dilution for evaluation of the carrier solvent as the reaction solvent (*11*, *12*). The diluted reaction segments would continuously flow through a reactor coil with precise control of flow rate, temperature, pressure, and residence time. Real-time analysis of the segments as they emerged from the reactor coil, via fractionation into UPLC-MS, would eliminate subsequent off-line analysis time.

Flow technology system configuration

Our system configuration is depicted in Fig. 1A. Given the potential air and moisture sensitivity of the chemistries under evaluation and that we would be handling the ligands and catalysts in solution, the reactor was assembled in a glovebox environment.

The system uses a modified high-performance liquid chromatography (HPLC) system with a well-plate autosampler to prepare the reaction segments following a user-defined injector program, which guides the accurate aspiration and injection of microliter volumes from up to 192 source vials. Through this methodology, we optimized the system such that it takes the autosampler 45 s to assemble a reaction segment from, in this instance, five components composed of reactants 1 and 2 (for example, aryl halide and aryl boronic

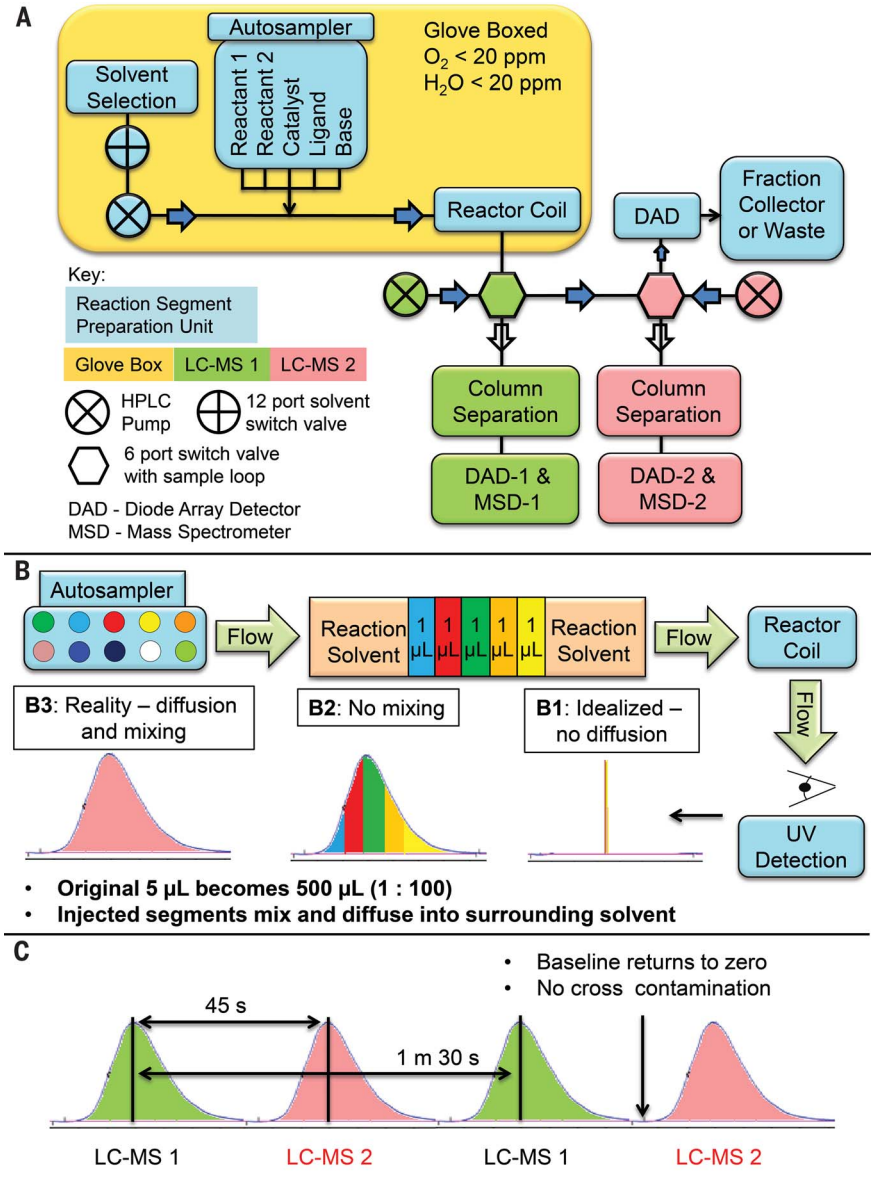


Fig. 1. Flow system setup and segment preparation. (**A**) Schematic depiction of the flow system. (**B**) Segment preparation and injection into flow stream showing potential mixing and diffusion outcomes, (B1) idealized–no diffusion, (B2) no mixing, and (B3) observed. (**C**) Portrayal of UV trace of the emerging reactions segments and fractionation into alternating LC-MS units.

acid in the case of a Suzuki-Miyaura coupling), catalyst, ligand, and base. The reaction segments are then injected into a flowing solvent stream (Fig. 1B), which is predetermined by the programmed method using the 12-port solvent selection value. The segment residence time in the temperature controlled reactor coil is determined by the pump flow rate.

To ensure maximal time efficiency, it is critical for the reaction samples to be analyzed as soon as they emerge from the reactor (*13*). To achieve this aim, two Agilent 1200 UPLC-MS instruments were positioned after the reactor outside of the glovebox so that while one was analyzing a reaction segment, the other was waiting to be triggered to analyze the next emerging segment (generated in 45-s intervals). In this manner, the three key dynamic components of the instrument (the two UPLC-MS instruments and the reaction segment preparation unit) are continuously working synergistically for maximum throughput. As a segment emerges from the reactor coil, it encounters a six-port switching valve, which directs it to the vacant LC-MS instrument for a detailed analysis (for clarity, segment 1 to LC-MS 1, segment 2 to LC-MS 2, etc.; Fig. 1C). The excess sample is directed through a diode array to enable visualization of the segment, and then either to a fraction collector or to waste, depending on the instrument's mode of operation.

Validation of four-component mixing in a model Suzuki-Miyaura reaction

With the general reactor design in hand, we first needed to assess the homogeneity of mixing of the various reaction components within the reaction segment, as well as the extent of diffusion of the segment itself into the surrounding carrier solvent, which is the key differentiating principle of this technology (*14*). Given its prevalent use in medicinal chemistry (*15*, *16*) as well as the extensive variables for optimization (ligands, Pd source, base), the Suzuki-Miyaura coupling between 6-bromoquinoline (**1b**) and the indazole boronic acid (**2a**) was chosen to validate the platform (Fig. 2A) (*17*). One of the challenges in evaluating homogeneity of mixing is that each of the reaction components will respond differently in the LC-MS, and the output trace will further be complicated by the extent of conversion to product. To counter this issue, we chose to make up the reaction component stock solutions of our model reaction (Fig. 2B) with inert internal standards such that the level of four of the five components could be accurately monitored. The internal standard used for each component was matched to the solubility, ensuring homogeneity in all cases. For the readily soluble components, ultraviolet (UV) active nonpolar solvents were used: **1b** dissolved in 1,3-diethylbenzene, $Pd(OAc)_2$ dissolved in 1,3,5-triethylbenzene, and PPh_3 dissolved in toluene. For **2a**, which is considerably less soluble, dimethylformamide (DMF) was used with 4,4′-di-*tert*-butylbiphenyl added as an internal standard. The fifth component, 1 M aqueous NaOH, was not directly monitored because of the difficulty in identifying an aqueous soluble inert internal standard. The molarities of the solutions were made such that a 1-µl addition of each (5 µl total) equates to a ratio of 1: 1: 2.5: 0.125: 0.0625 (**1a**: **2b**: base: ligand: Pd). With methanol as the carrier solvent, the segments were run with a flow rate of 1 ml/min through a Hastelloy coil heated to 100°C with a residence time of 1 min.

In this experiment, reaction segments of increasing volume were created with 1, 2, 4, 8, and 16 µl of each component (five components thus totaling 5, 10, 20, 40, and 80 µl) to assess the degree of mixing, the extent of diffusion, and the scalability of the reaction technology across a wide range of reaction volumes. To make up the segment, the five components were aspirated in series and then allowed to diffuse together, thus presenting a key differentiation with typical segmented flow systems in which diffusion is inhibited through use of a spacer, typically an inert gas or a perfluorinated solvent (*11*). Once the segments passed through the reactor coil, the output was split into a 96-well plate in 40-µl fractions. Each of the fractions was then analyzed via an off-line LC-MS by a standard method that enabled separation of each of the four internal standards mentioned previously. As shown in Fig. 2B, the ratio of each of the internal standards is approximately equivalent throughout the segment, demonstrating homogeneous diffusion of the reaction components into the

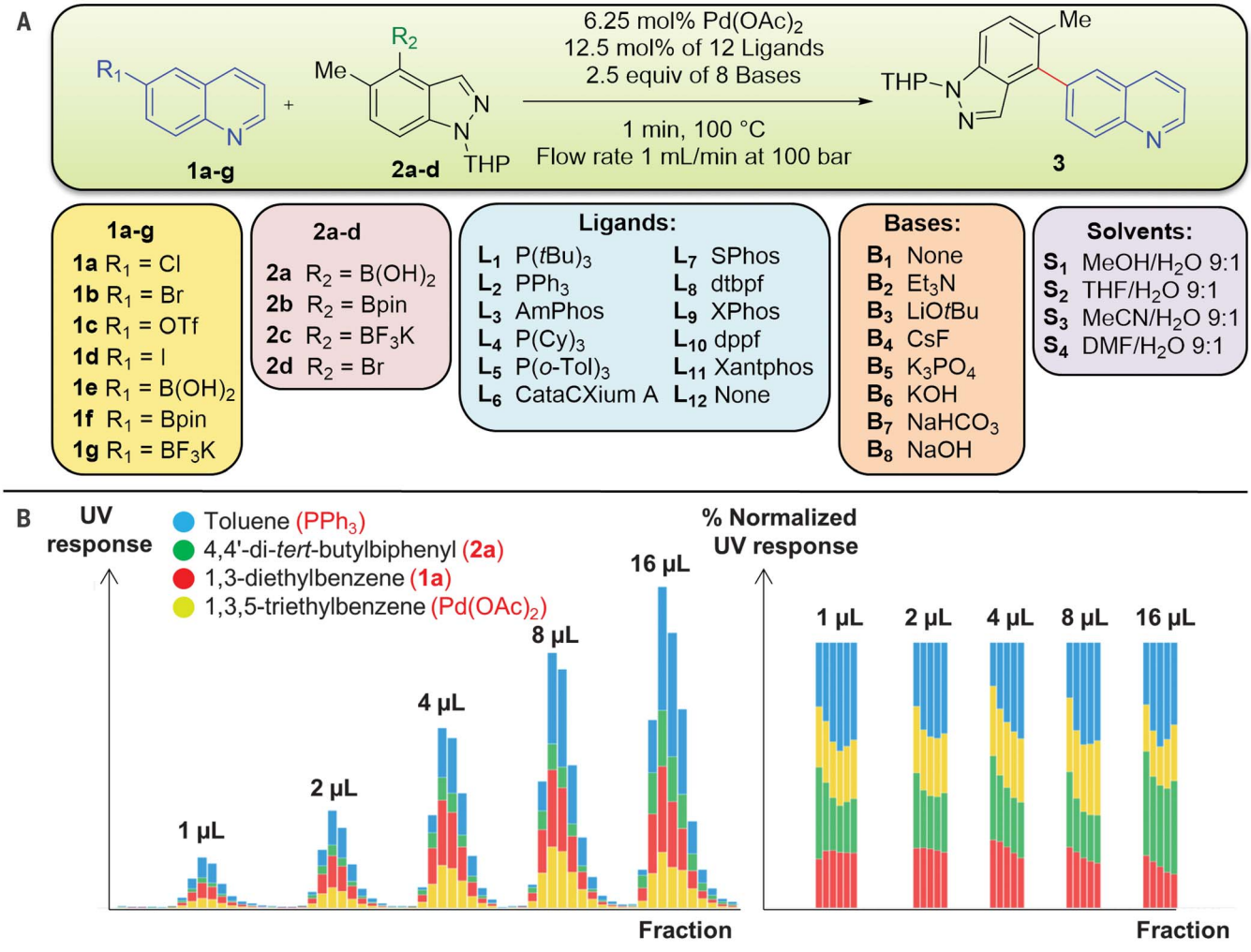


Fig. 2. Model Suzuki-Miyaura cross-coupling. (**A**) Coupling evaluates electrophiles **1a–1d**/**2d**, with **2a–2c**/**1e–1g** evaluated across a matrix of 11 ligands (plus one blank) × 7 bases (plus one blank) × 4 solvents. (**B**) UV analysis of internal standards (see legend) in the fractionated reaction segment derived from increasing volumes of components confirming adequate mixing.

surrounding carrier solvent, over a wide range of injection volumes. This result confirms that consistent reaction component stoichiometries can be generated throughout the segment and that larger segments can be injected to directly scale up screening results to produce meaningful quantities of material for biological evaluation. Additionally, these results show substantial homogeneous diffusion into the carrier solvent, which allows solvents to be screened with the same stock solutions in a continuous manner, a strategy that has not previously been achieved in the flow chemistry paradigm.

Rapid material-sparing screening of 5760 reactions

Judicious selection of both coupling partners plays a critical role, in terms of both the reactivity and economics of a Suzuki-Miyaura coupling (*18*). We therefore expanded upon our model transformation to execute on this concept, evaluating the coupling of electrophiles **1a** to **1d** with the nucleophiles **2a** to **2c** as well as the reverse combination **2d** with **1e** to **1g** across a matrix of 11 ligands (plus one blank) × 7 bases (plus one blank) × 4 solvents (Fig. 2A). Selection of the ligands was based on a series of factors, specifically (i) performance in previous internal batch screens; (ii) an internal principal component analysis of ligand property space (*19*, *20*); (iii) coverage of the main ligand classes from the literature (*21*); and (iv) commercial availability. The bases selected span a range of both organic and inorganic commonly used in this transformation, and solvents were chosen to display a range of polarities and dipole moments and the presence or absence of hydrogen bond donors (*22*). All reactions used $Pd(OAc)_2$ as precatalyst. Given that most Suzuki-Miyaura reactions use water as a cosolvent to solubilize the inorganic base employed and promote the formation of the boronate complexes involved in the transmetallation (*23*), the pump was set to provide a 9:1 solvent/water ratio (*24*). Overall, screening the complete set of variables would provide data for a total of 5760 reactions.

One important goal in our initial experimentation was to confirm that differences in reactivity between various reactions are due to the bulk carrier solvent and not the solvent in which the reagent stock solutions were made. Given the common nature of the solvents used to make up the stock solutions, we can infer that if there were a change in the reactivity during the solvent screen, it would be attributable to the carrier solvent mediating the reaction given the 100:1 dilution factor.

All relevant stock solutions were prepared under an inert atmosphere with the system set up as described previously for the validation of mixing experiments with the two UPLC-MS systems operating in tandem for analysis. The screen was run by making up successive reaction segments at 45-s intervals with 1 µl of each component stock solution. The stock solutions were

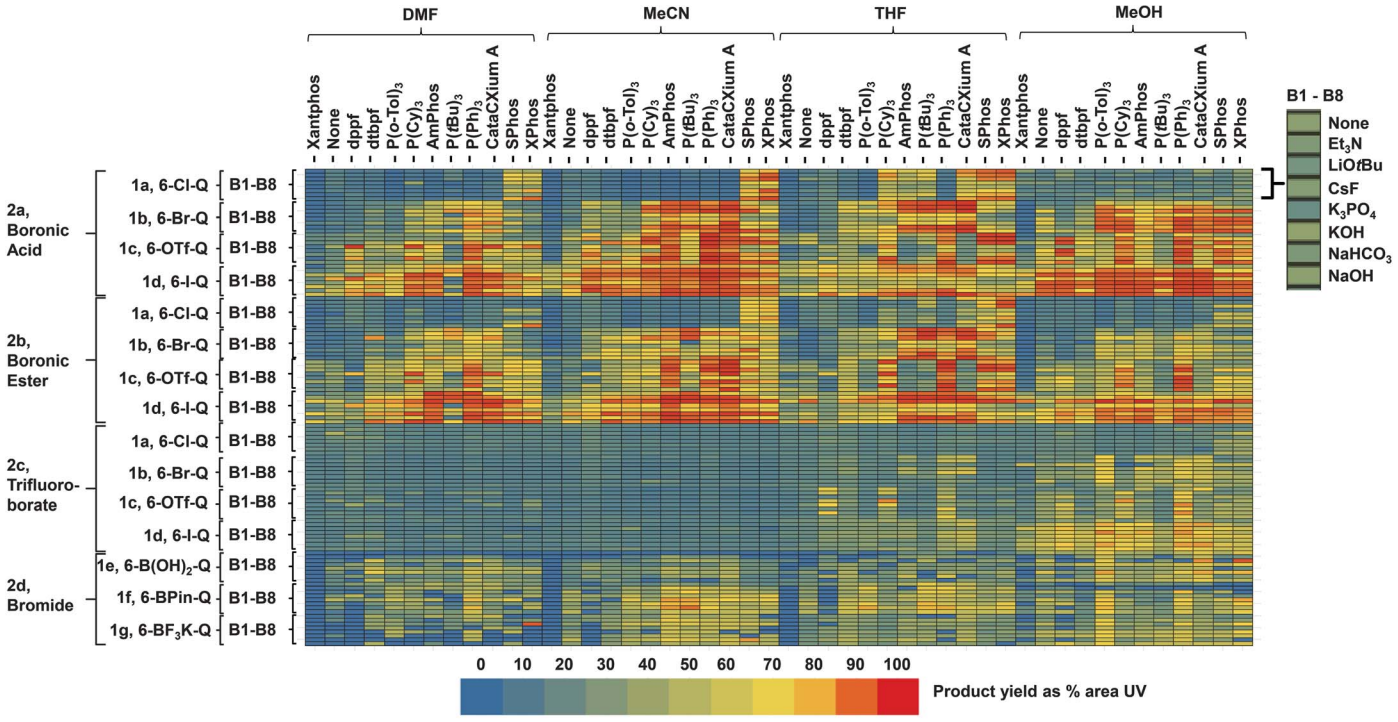


Fig. 3. Complete heatmap visualization. A total of 5760 reactions of **1a–1d** with **2a–2c**, and the reaction of **2d** with **1e–1g** evaluated across a matrix of 11 ligands (plus one blank) × 7 bases (plus one blank) × 4 solvents.

made up at molarity that equated to stoichiometries of 1: 1: 2.5: 0.125: 0.0625 (reactant 1: reactant 2: base: ligand: Pd). Each reaction was run with just 0.4 μmol, which in the case of **1a** equates to 65 μg each or 98 mg per 1500 reactions within 24 hours. The segments were then injected into the solvent stream flowing at 1 ml/min and 100 bar of pressure. This short reaction time is necessary to expedite throughput, and we anticipated that regardless of the absolute conversion to product, the relative conversions would allow the judicious selection of reaction conditions to be investigated for scale-up. Once a run was initiated, the system operated in a fully automated manner, at a rate of 1500 reactions per 24-hour time period. In contrast, a typical batch screen of 192 reactions in our laboratories runs for 18 hours (*25*).

To expedite data analysis, we used the Agilent Chemstation software in real time to identify key peaks in the LC-MS trace, followed by off-line refinement of the data through the iChemExplorer software before export and visualization with Spotfire, the latter allowing a large degree of flexibility in the identification of key reactivity trends. The overall process for data manipulation took approximately an hour for 1500 reactions. The complete set of data for the 5760 reactions is presented in the heatmap shown in Fig. 3.

Despite the large amount of data presented therein, it is possible to determine some high-level reactivity trends from this visualization. First, the 6-chloroquinoline electrophile **1a** is clearly the worst substrate, although there are conditions identified showing good reactivity (specifically with ligands such as XPhos and SPhos) (*26*, *27*). The relative reactivity of the quinoline cores can be more explicitly seen by the box plot shown in Fig. 4A, which compares the yields of these four electrophiles (**1a** to **1d**) in the reactions with both the boronic acid (**2a**), BPin (**2b**), and BF_3K (**2c**) derivatives. Furthermore, it is clear that a higher density of conditions provides high conversion when the indazole is used as the nucleophilic partner rather than as the electrophile.

Second, comparison of the boron sources shows that the BF_3K derivative **2c** is inferior to both the boronic acid **2a** and ester **2b**, which may reflect that the 1-min residence time is not sufficient for the BF_3K (**2c**) derivative to hydrolyze to the boronic acid (**2a**) for coupling (*23*). However, in the cases where the BF_3K (**2c**) has shown reactivity, the use of MeOH (or to a lesser degree tetrahydrofuran (THF) as the solvent is key (Fig. 4A), thus validating the concept of identifying solvent reactivity trends through variation of the bulk carrier solvent.

Focusing specifically on the reaction of the boronic acid **2a** with the four quinoline-based electrophiles **1a** to **1d**, again we can clearly see that chloroquinoline **1a** performs most poorly (Fig. 4B). However, it is also possible to determine further trends specifically for the other electrophiles. We can observe that in general, for bromoquinoline **1b**, methanol appears to be the superior solvent. We can also see that in general, Xantphos is a poor choice of ligand for the reaction, whereas PPh_3 can give high conversions to the desired product for all but the Cl-quinoline, particularly in MeOH and MeCN (*28*).

Finally, it is possible to filter the conditions, and cluster by those that provide a conversion of greater than 85% for all of the electrophiles screened. This result, shown in Fig. 5, demonstrates that from the 384 potential conditions evaluated, three will work well independent of the selection of the electrophile, and these are all based either on X-Phos or S-Phos and use MeCN as the solvent. This finding is key for identifying conditions that are most amenable for parallel synthesis—for example, wherein one seeks reactivity across the broadest substrate scope.

To demonstrate the capability of the instrument for providing milligram quantities of material through the processing of multiple segments, we chose a representative set of reaction conditions from the screening results discussed. Using these conditions, we programmed the autosampler to inject 100 consecutive segments consisting of 8 μl per reaction component (3.2 μmol per segment), which would be able to provide a maximum of 110 mg of desired material in ~75 min. The combined segments were collected and evaporated, and the product isolated in 59% yield (65 mg) after purification by column chromatography (Fig. 6A). The reaction conditions identified were translated to further scale-up in continuous flow with a commercially available Vaportec R series reactor, with 42% of the desired product being obtained by means of the standard two-pump setup (*29*).

We next aimed to demonstrate reproducibility and direct translation to traditional batch reaction operations. With no further optimization or experimentation in batch, the flow-identified conditions were altered prospectively only to make the overall reaction more concentrated (0.10 M) and carried out with typical Schlenk line techniques (*30*). Gratifyingly, the reaction performed well in batch and gave an isolated yield of 79% for the desired coupling product, on time scales similar to that for the flow optimization work (Fig. 6A). To further validate the results obtained from the screen, several additional permutations of reaction conditions including solvents, ligands, and coupling partners were run in a batch manner. Two positive controls gave isolated yields of 67 and 93%, respectively, whereas two negative controls gave either <10% conversion or no trace of product after reaction overnight (see supplementary materials for detailed experimental conditions for control scale-up experiments 9 to 12). Although these further results cannot totally rule out the possibility of "false positives" or "false negatives" in the data set, they build further confidence in the validity and reproducibility of the experimental outcomes provided by this method.

To further demonstrate the utility of the reaction platform for optimizing a particularly challenging Suzuki-Miyaura coupling, we did an analysis of recently conducted Suzuki-Miyaura coupling parallel synthesis libraries [parallel medicinal chemistry (PMC)] within ongoing Pfizer discovery projects. The bromo-oxindole, **5**, rose to the top as an electrophile that was present in the design of 13 libraries (*31*), reflecting its desirable properties from a medicinal chemistry perspective, but failed to provide the desired product in all cases. A screen of conditions, which at 0.4-μmol scale per reaction equates to 50 mg of **5** per 576 reactions in 8 hours for the coupling with boronic ester **4**, showed a stark disparity in the efficiency of conditions for this coupling (Fig. 6B), with only a few providing moderate (45 to 65%) conversion to the desired product during the 1-min reaction time screened, as indicated on the heatmap (Fig. 6C). These results suggested that CataCXium A (*32*) was a uniquely effective catalyst with THF/H_2O as the preferred solvent and MeOH/H_2O as a potential alternative (*33*). Et_3N was preferred as the base, with NaOH, CsF, and K_3PO_4 emerging as possible inorganic alternatives. Finally, the conditions typically used for our PMC synthesis [$Pd(OAc)_2$, $P(Cy)_3$, THF/H_2O] completely failed in the flow screen, thus supporting the original findings.

Direct translation of these optimal conditions to batch, with the exception of reaction concentration and elongated reaction time, afforded high conversion and isolated yield of the coupling product, which was not observed in the PMC efforts (Fig. 6B).

Outlook

The platform described in this study provides a notable advance in the ability to screen reaction reagents, solvents, and conditions in a flow-based

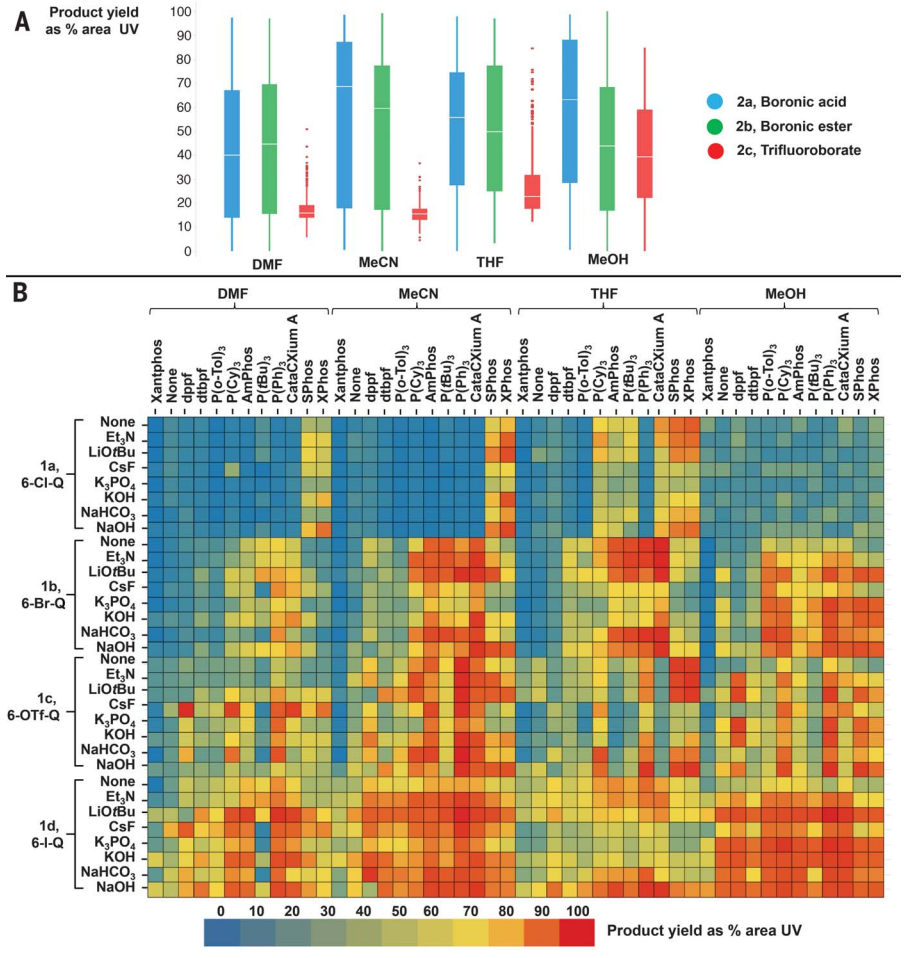
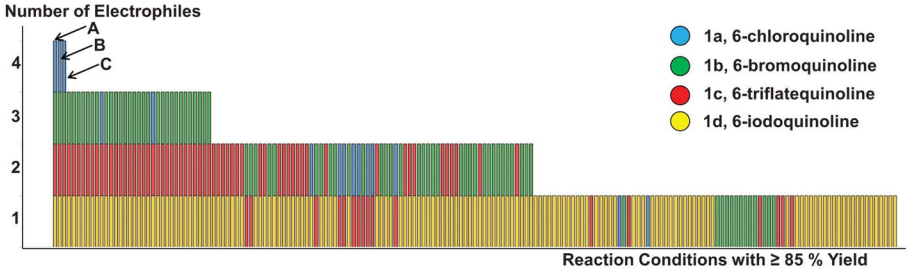


Fig. 4. Data analysis of model Suzuki-Miyaura cross-coupling. (**A**) Box plot comparison of quinoline electrophiles **1a–1d** in reactions with boronic acid **2a** and boronic ester derivatives, **2b** and **2c**, across different solvents. (**B**) Heatmap of coupling of quinoline electrophiles **1a–1d** with **2a** evaluated across a matrix of 11 ligands (plus one blank) × 7 bases (plus one blank) × 4 solvents.



Reaction Condition	Solvent	Base	Ligand	6-Cl-Q (1a)	6-Br-Q (1b)	6-OTf-Q (1c)	6-I-Q (1d)
A	MeCN	LiO*t*Bu	SPhos	85	92	91	93
B	MeCN	NaOH	SPhos	86	93	91	88
C	MeCN	NaOH	XPhos	93	90	92	86

Fig. 5. Suzuki-Miyaura cross-coupling—successful conditions. Conditions with ≥85% yield for all quinoline-based electrophiles **1a–1d** with **2a**. This indicates that 181 conditions work for one electrophile, 103 conditions work for two electrophiles, and 33 conditions work for three electrophiles, whereas only 3 sets of reaction conditions (specified in the table) work for all four electrophiles.

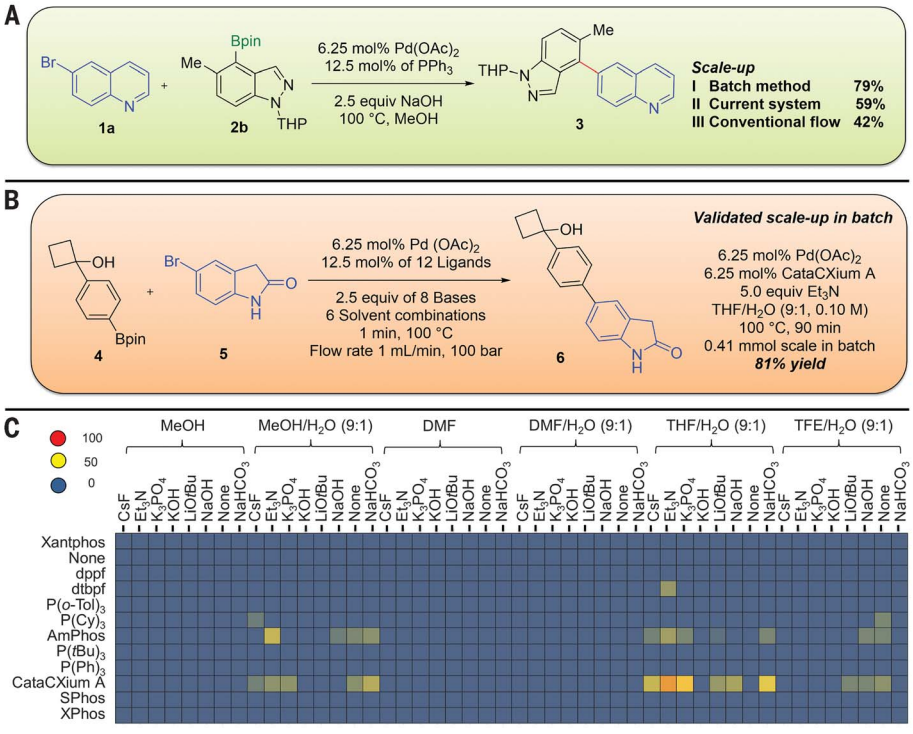


Fig. 6. Scale-up validation and PMC optimization. (**A**) Scale-up validation of reaction conditions in both batch and flow for the Suzuki-Miyaura cross-coupling of **1b** with **2b**. (**B**) PMC optimization via flow-screening and validation in batch for the coupling of **4** with **5** evaluating a matrix of 11 ligands (plus one blank) × 7 bases (plus one blank) × 6 solvents. (**C**) Complete heatmap visualization of the 576 reactions.

manner. Although it is typically difficult to directly compare batch and flow platforms based on the limited information provided regarding setup, the ability described here to run more than 1500 reactions in a 24-hour period with high-resolution information-rich reaction analysis benchmarks this methodology favorably against plate-based techniques. In addition, there are a number of advantages, beyond the generation of real-time analytical data, associated with running the reactions in a flow paradigm, including avoiding solvent evaporation, improved mixing, and uniform heating. The system at present is geared toward the analysis of homogeneous reactions, and thus issues arising from inefficient mixing of either heterogeneous or biphasic reaction systems present a gap, which remains to be addressed in the current technology.

REFERENCES AND NOTES

1. M. B. Plutschack, B. Pieber, K. Gilmore, P. H. Seeberger, *Chem. Rev.* **117**, 11796–11893 (2017).
2. D. E. Fitzpatrick, C. Battilocchio, S. V. Ley, *ACS Cent Sci* **2**, 131–138 (2016).
3. *New Synthetic Technologies in Medicinal Chemistry*, E. Farrant, Ed. (RSC, Cambridge, 2012), pp. 1–164.
4. J. A. DiMasi, H. G. Grabowski, R. W. Hansen, *J. Health Econ.* **47**, 20–33 (2016).
5. T. Cernak *et al.*, *J. Med. Chem.* **60**, 3594–3605 (2017).
6. A. Buitrago Santanilla *et al.*, *Science* **347**, 49–53 (2014).
7. D. K. B. Mohamed, X. Yu, J. Li, J. Wu, *Tetrahedron Lett.* **57**, 3965–3977 (2016).
8. B. J. Reizman, K. F. Jensen, *Acc. Chem. Res.* **49**, 1786–1796 (2016).
9. A. Günther, K. F. Jensen, *Lab Chip* **6**, 1487–1503 (2006).
10. B. J. Reizman, K. F. Jensen, *Chem. Commun. (Camb.)* **51**, 13290–13293 (2015).
11. N. Hawbaker, E. Wittgrove, B. Christensen, N. Sach, D. G. Blackmond, *Org. Process Res. Dev.* **20**, 465–473 (2015).
12. Mixing has been reported as a potential issue in heterogeneous reactions (*34*), although these are outside the scope of the current work as the organic solvents in this study were judiciously selected to provide homogeneous solutions when mixed in a 9/1 ratio with water.
13. N. Holmes *et al.*, *Reaction Chemistry & Engineering* **1**, 96–100 (2016).
14. A. Günther, M. Jhunjhunwala, M. Thalmann, M. A. Schmidt, K. F. Jensen, *Langmuir* **21**, 1547–1555 (2005).
15. D. G. Brown, J. Boström, *J. Med. Chem.* **59**, 4443–4458 (2016).
16. S. D. Roughley, A. M. Jordan, *J. Med. Chem.* **54**, 3451–3479 (2011).
17. J. Magano, J. R. Dunetz, *Chem. Rev.* **111**, 2177–2250 (2011).
18. A. J. J. Lennox, G. C. Lloyd-Jones, *Chem. Soc. Rev.* **43**, 412–443 (2014).
19. J. Jover *et al.*, *Organometallics* **31**, 5302–5306 (2012).
20. J. Jover *et al.*, *Organometallics* **29**, 6245–6258 (2010).
21. P. G. Gildner, T. J. Colacot, *Organometallics* **34**, 5497–5508 (2015).
22. P. M. Murray *et al.*, *Org. Biomol. Chem.* **14**, 2373–2384 (2016).
23. A. J. J. Lennox, G. C. Lloyd-Jones, *Angew. Chem. Int. Ed.* **52**, 7362–7370 (2013).
24. Although water was incorporated into the Suzuki-Miyaura reaction described herein, further experimentation has demonstrated that the system performs equally well for nonaqueous reaction screens.
25. For an example of a typical reaction screen from our laboratories, see (*35*).
26. X. Huang *et al.*, *J. Am. Chem. Soc.* **125**, 6653–6655 (2003).
27. T. E. Barder, S. D. Walker, J. R. Martinelli, S. L. Buchwald, *J. Am. Chem. Soc.* **127**, 4685–4696 (2005).
28. P. C. J. Kamer, P. W. N. M. van Leeuwen, J. N. H. Reek, *Acc. Chem. Res.* **34**, 895–904 (2001).
29. With regard to the modest yields obtained with the Vaportec system, no optimization work was undertaken in terms of combination and stoichiometry of the reagents or residence time in the reactor. In addition, using a two-pump setup necessitated making up mixed solutions of several reaction components before mixing. The major observed by-product in these experiments was homocoupling of the quinoline, **1b**.
30. A potential question arises as to whether the solvent used to make up the stock solution can influence the reaction. However, as demonstrated by Jouyban *et al.* (*36*), the impact of a mixed solvent in terms of a dielectric constant is a weighted average of the mixed components. As such, the effect of injecting a different solvent into a 9/1 organic/aqueous mixture that is diluted out 1:100 will largely be negated so long as it is inert to the chemistry being screened. The validity of this argument is borne out by the fact that the scaled batch experiments work in a similar manner to the flow experiments even without the solvents used to make up the stock solutions. However, although no effect of the small amount of stock solvent present in the reaction system has been observed thus far, we cannot preclude its potential involvement in all transformation screening.
31. The monomer **5** was used in 13 in-house library campaigns all involving Pd-mediated couplings (9 were Suzuki-Miyaura couplings whereas the remainder were Buchwald-Hartwig reactions). In none of the examples evaluated did **5** lead to any of the desired product in the library matrix.
32. A. Zapf, A. Ehrentraut, M. Beller, *Angew. Chem. Int. Ed.* **39**, 4153–4155 (2000).
33. These results reinforce the efficiency of the mixing within the system, including the aqueous component. In this experiment, two systems are evaluated in which "pure" organic solvents (DMF, MeOH) have been utilized, with no reactions performing well. This is to be expected as the experiment uses the BPin derivative, which to react requires hydrolysis to the $B(OH)_2$ derivative (*22*). If sufficient mixing did not occur with the aqueous-based systems then poor reactivity would be expected for all reactions.
34. J. R. Naber, S. L. Buchwald, *Angew. Chem. Int. Ed.* **49**, 9469–9474 (2010).
35. Q. Huang *et al.*, *Org. Process Res. Dev.* **15**, 556–564 (2011).
36. A. Jouyban, S. Soltanpour, H.-K. Chan, *Int. J. Pharm.* **269**, 353–360 (2004).

ACKNOWLEDGMENTS

We thank L. Bernier, J. Braganza, M. Collins, K. Dress, J. Lafontaine, G. Ng, U. Reilly, D. Richter, T. Long, G. Steeno, C. Subramanyam, and D. Truong for helpful discussions. D. P. was supported by postdoctoral research fellowship from Pfizer. Additional data supporting the conclusion are available in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6374/429/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S24
Tables S1 to S3
References (*37*–*40*)
Data File S1

8 September 2017; accepted 13 December 2017
10.1126/science.aap9112

REPORT

ORGANIC CHEMISTRY

Synthesis of partially and fully fused polyaromatics by annulative chlorophenylene dimerization

Yoshito Koga,[1] Takeshi Kaneda,[1] Yutaro Saito,[1] Kei Murakami,[1,2]* Kenichiro Itami[1,2,3]*

Since the discovery by Ullmann and Bielecki in 1901, reductive dimerization (or homocoupling) of aryl halides has been extensively exploited for the generation of a range of biaryl-based functional molecules. In contrast to the single-point connection in these products, edge-sharing fused aromatic systems have not generally been accessible from simple aryl halides via annulation cascades. Here we report a single-step synthesis of fused aromatics with a triphenylene core by the palladium-catalyzed annulative dimerization of structurally and functionally diverse chlorophenylenes through double carbon-hydrogen bond activation. The partially fused polyaromatics can be transformed into fully fused, small graphene nanoribbons, which are otherwise difficult to synthesize. This simple, yet powerful, method allows access to functional π-systems of interest in optoelectronics research.

The reductive dimerization (or homocoupling) of aryl halides (*1*) has been extensively exploited in organic synthesis since Ullmann and Bielecki's discovery of the copper-mediated dimerization of aryl halides in 1901 (*2*) (Fig. 1A). This reaction has proved to be particularly useful in the synthesis of biaryl-based functional molecules, including pharmaceuticals, biologically active natural products, optoelectronic π-conjugated materials, and polymers. Although this textbook reaction is a powerful technique for the single-point connection of two aromatic nuclei, methods to access fused aromatic systems from simple aryl halides, via annulation cascades, have not been developed (Fig. 1A). Such fused polycyclic aromatics represent an emerging class of π-conjugated molecules in optoelectronic devices, nanographene materials (*3*), ultrashort carbon nanotubes (*4*), and carbon nanobelts (*5*). Their optoelectronic properties are susceptible to structural perturbations based on shape, width, edge topology, and degree of π-extension. Thus, a method to synthesize fused aromatics in a bottom-up fashion with atom-by-atom precision is in high demand (*3*–*7*). In particular, methods that involve C–H functionalization (*8*, *9*), a direct molecular activation-transformation technology, are attractive not only to streamline overall synthesis but also to unlock opportunities for markedly different reactivity and selectivity. On the basis of our single-step π-extension strategy for making new nanocarbon molecules by C–H functionalization (*10*, *11*), we envisioned that ortho-C–H bonds of aryl halides might be activated for annulative dimerization processes to access π-extended, fused polyaromatic systems (Fig. 1A).

Here we report a single-step synthesis of triphenylene-cored, fused aromatics by the palladium-catalyzed annulative dimerization of structurally and functionally diverse chlorophenylenes (*12*, *13*) through double C–H activation (Fig. 1B). The overall reaction is redox-neutral, such that stoichiometric reductant is not required. Chlorinated aromatic reagents are readily available. This simple, yet powerful, dimerization allows for the fusion of two functional aromatic nuclei to directly access triphenylene-cored π-systems that are known as privileged structures for materials in organic light-emitting diodes (OLEDs) (*14*–*17*). Moreover, the thus-generated partially fused aromatics can be transformed into fully fused, small graphene nanoribbons, which are otherwise difficult to synthesize.

[1]Graduate School of Science, Nagoya University, Chikusa, Nagoya 464-8602, Japan. [2]Institute of Transformative Bio-Molecules (WPI-ITbM), Nagoya University, Chikusa, Nagoya 464-8602, Japan. [3]Japan Science and Technology Agency–Exploratory Research for Advanced Technology (JST-ERATO), Itami Molecular Nanocarbon Project, Nagoya University, Chikusa, Nagoya 464-8602, Japan.
*Corresponding author. Email: itami@chem.nagoya-u.ac.jp (K.I.); murakami@chem.nagoya-u.ac.jp (K.M.)

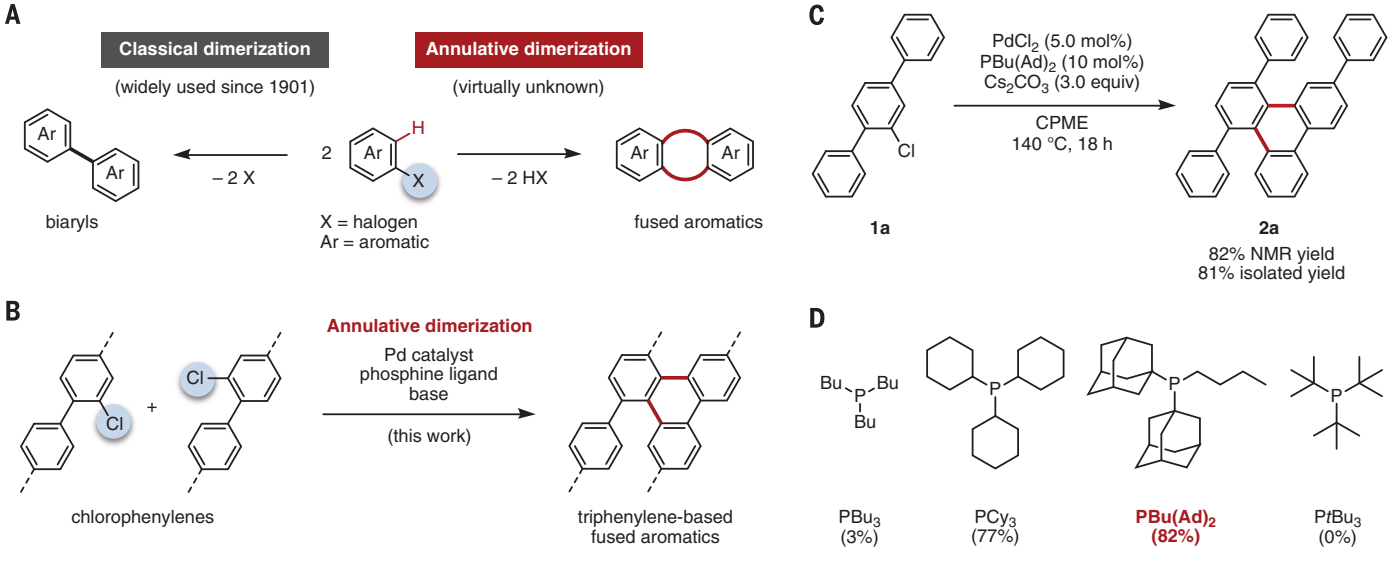
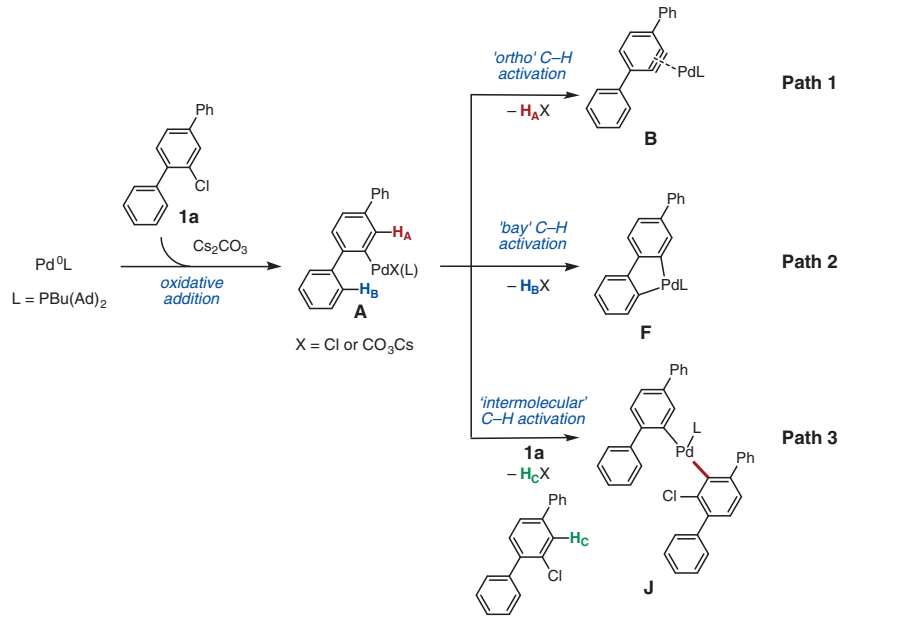


Fig. 1. Potential of annulative dimerization of aryl halides. (**A**) Classical reductive dimerization and annulative dimerization. (**B**) Pd-catalyzed annulative dimerization of chlorophenylenes to give triphenylene-based fused aromatics. (**C**) Optimized reaction conditions: **1a** (1.0 equivalent), $PdCl_2$ (5.0 mol %), $PBu(Ad)_2$ (10 mol %), Cs_2CO_3 (3.0 equivalents), CPME, 140°C, 18 hours. (**D**) Effect of phosphine ligands. Conditions: **1a** (1.0 equivalent), $PdCl_2$ (5.0 mol %), ligand (10 mol %), Cs_2CO_3 (3.0 equivalents), CPME, 140°C, 18 hours. ^1H NMR yields of **2a** are given in parentheses. Red bonds indicates those that are newly formed.

We began our campaign by establishing the conditions for the annulative dimerization reaction using 2′-chloro-1, 1′:4′, 1″-terphenyl (**1a**) as a representative chlorinated oligophenylene. Through extensive screening of various catalysts and additives, we identified the optimized reaction conditions: Treatment of **1a** (1.0 equivalent) with $PdCl_2$ [5.0 mole % (mol %)], $PBu(Ad)_2$ (10 mol %; Bu, *n*-butyl; Ad, 1-adamantyl), and cesium carbonate (3.0 equivalents) in cyclopentyl methyl ether (CPME) at 140°C for 18 hours afforded the annulative dimerization product **2a** in 81% isolated yield (Fig. 1C). The structure of **2a** was unambiguously confirmed by x-ray crystallographic analysis. The ligand had a critical effect on the reaction efficiency; representative results are shown in Fig. 1D. The use of less bulky PBu_3 hindered the reaction. Moderately large trialkylphosphines PCy_3 (Cy, cyclohexyl) and $PBu(Ad)_2$ afforded the best results (77 and 82%), whereas much bulkier $PtBu_3$ (*t*Bu, *tert*-butyl) did not facilitate the dimerization. Other phosphine ligands P(*o*-tol)$_3$ (*o*-tol, *o*-$CH_3C_6H_4$) and 1,2-bis(dicyclohexylphosphino)ethane, as well as N-heterocyclic carbene ligands, were not effective for the reaction.



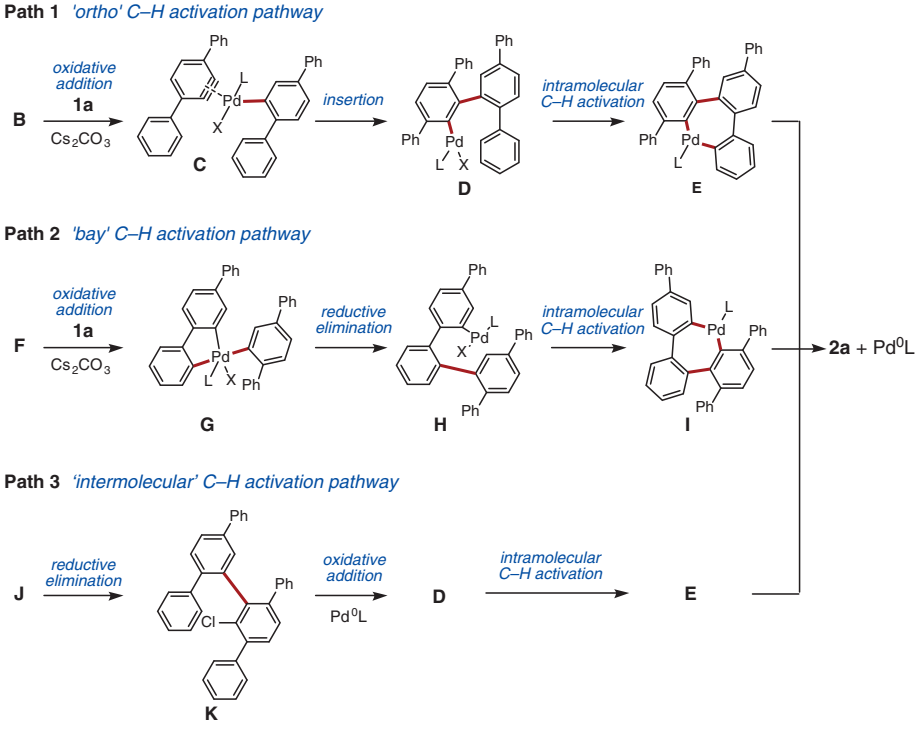


Fig. 2. Possible reaction pathways. Three pathways are shown: path 1, ortho C–H activation pathway; path 2, bay C–H activation pathway; and path 3, intermolecular C–H activation pathway. Ph, phenyl; L, $PBu(Ad)_2$; X, Cl or CO_3Cs.

Although the exact mechanism of the annulative dimerization remains unclear, our current assumption is shown in Fig. 2. An in situ–generated, phosphine-bound Pd(0) species Pd^0L [L, $PBu(Ad)_2$] undergoes the first oxidative addition with chloroterphenyl **1a** to give arylpalladium **A**. The resulting arylpalladium **A** can abstract three hydrogen atoms via path 1, intramolecular H_A (red); path 2, intramolecular H_B (blue); or path 3, intermolecular H_C (green). Path 1 initiates from intramolecular ortho-C–H activation (*18*) from **A** to give Pd-aryne **B** (*19*). Oxidative addition of **1a** to **B** provides **C**, which undergoes an insertion (carbopalladation) across the aryne moiety to give arylpalladium **D**. Subsequent intramolecular C–H palladation affords seven-membered palladacycle **E**, which then undergoes reductive elimination to provide triphenylene product **2a** and Pd^0L, completing the catalytic cycle. As an alternative to the Pd-aryne pathway (path 1), the Pd(IV) pathway (path 2) might also be possible (*20*). For example, intramolecular C–H activation of **A** at the baylike region affords palladacycle intermediate **F**. Subsequent oxidative addition of **1a** generates Pd(IV) intermediate **G**, which then undergoes biaryl-forming reductive elimination to yield arylpalladium **H**. Intramolecular C–H activation affords **I**, which finally gives **2a** and Pd^0L. Path *3* involves intermolecular C–H activation of **A** with **1a**, whereby the C–H bond ortho to the chlorine atom is activated, to give intermediate **J**, which then provides **K**. Subsequent intramolecular C–H arylation of **K** affords **2a** through intermediates **D** and **E**.

With the optimal conditions for the annulative dimerization in hand, the scope of the reaction was investigated with a range of structurally and functionally diverse chlorophenylenes (Fig. 3A). Terphenyl **1b** with a *t*Bu group was converted into the corresponding triphenylene derivative **2b** in 77% isolated yield. Terphenyls with an electron-donating methoxy, trifluoromethoxy, methylsulfanyl, or electron-withdrawing trifluoromethyl group (**1c** to **1f**) were smoothly converted into **2c** to **2f** in good to high yields (85, 68, 49, and 64%, respectively). The methylsulfanyl group can be transformed through metal-catalyzed cross-coupling reactions as a halogen equivalent. Silyl groups, which can be used for further derivatizations, were also tolerated, and **2g** was obtained in 66% yield. In the reaction of **1h**, the methoxycarbonyl group remained intact to give **2h**. A small amount of isomer **2h′** was also generated, presumably because the ester group increased the acidity of the C–H bond on the arene to induce an unwanted palladium migration through the palladacycle intermediate (*21*). The reaction of **1i** afforded carbazole-containing triphenylene **2i** that could have potential as a hole-transporting material. *m*-Methyl– or *m*-phenyl–substituted **1j** and **1k** reacted smoothly to give **2j** and **2k** in 85 and 58% yields, respectively. Chlorobenzene derivatives with two naphthalene rings (**1l** and **1m**) dimerized to give fused π-extended systems **2l**

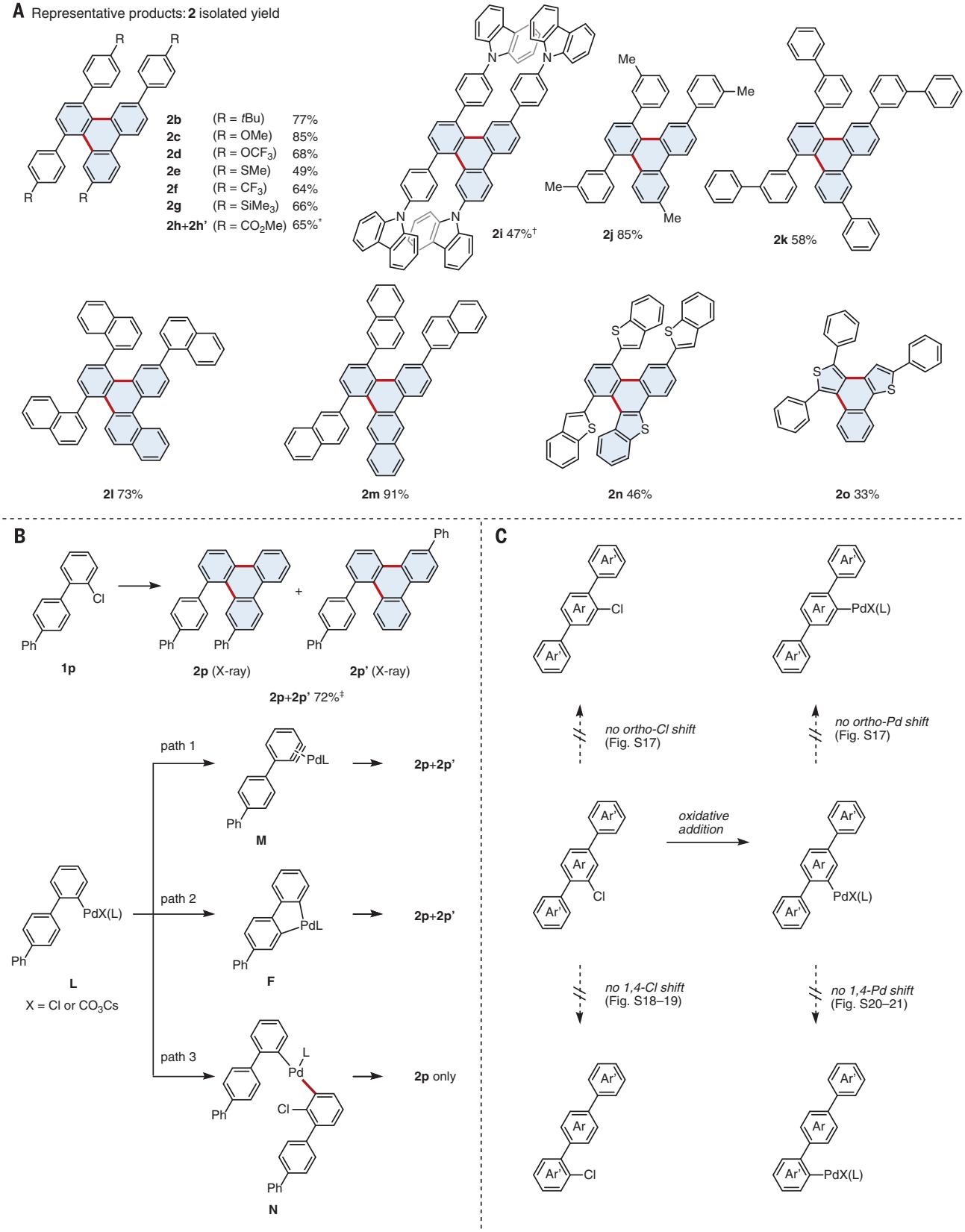


Fig. 3. Annulative dimerization with various substrates. Optimized reaction conditions: **1** (1.0 equivalent), $PdCl_2$ (5.0 mol %), $PBu(Ad)_2$ (10 mol %), Cs_2CO_3 (3.0 equivalents), CPME, 140°C, 18 hours. (**A**) Scope to examine functional group compatibility and structural diversity. *Ratio of **2h** to **2h′** is 3:1. For the structure of **2h′**, see supplementary materials. †CsF (3.0 equivalents) was used instead of Cs_2CO_3. Me, methyl. (**B**) Reaction of **1p**. ‡Ratio of **2p** to **2p′** is 3:1. (**C**) No chloride or palladium migration occur. L, $PBu(Ad)_2$; Ar, aromatic; X, Cl or CO_3Cs.

and **2m** in 73 and 91% yields with virtually complete regioselectivity. No further dehydrocyclization was observed. Benzothiophene-substituted derivative **1n** also reacted to give thiophene-benzene fused system **2n**. Chlorinated thiophene derivative **1o** also underwent annulative dimerization to give **2o** in 33% yield.

By applying various chloroarene substrates to the present annulative dimerization, mechanistic insights were also obtained. An overview of possible reaction pathways and the mechanistic experiments are provided in the supplementary materials. For example, chloroterphenyl **1p** is useful as a mechanistic probe (Fig. 3B). The reaction of **1p** resulted in the formation of a mixture of **2p** and **2p′**, structures of which were confirmed by x-ray crystallographic analysis. This shows that either path 1 or 2 was dominant and that the mixture of **2p** and **2p′** was generated through Pd-aryne intermediate **M** or palladacycle intermediate **F** (Fig. 3B). Intermediate **M** or **F** then reacted with **1p** to give two regioisomers. On the other hand, if path 3 had been dominant, the reaction would have resulted in the exclusive formation of **2p** via intermediate **N**. Importantly, this discussion is based on the hypothesis that no chloride or palladium migration occurs in the substrate and/or intermediates (Fig. 3C). From extensive experiments, we confirmed that such chloride or palladium migration is not taking place under the present reaction conditions (figs. S17 to S21). Taken together, we conclude that the intermolecular C–H activation pathway (path 3) is unlikely. Because both possible pathways, paths 1 and 2, involve previously unappreciated elementary reactions [the formation of Pd-aryne via ortho-C–H activation of arylpalladium or the oxidative addition of an aryl chloride to a palladacycle (*22*)], the mechanism of the present reaction is of great interest in its own right.

The construction of polycyclic aromatics has garnered considerable attention owing to their applicability in a range of functional materials. It is particularly important to synthesize these materials in a controlled fashion, because their structure profoundly affects their properties. In this study, all triphenylene-cored, π-extended compounds **2** display blue fluorescence with reasonably sharp spectral widths. The photophysical properties (ultraviolet through visible absorption spectra, fluorescence spectra, and fluorescence quantum yields) of **2** are provided in figs. S30 to S48. Coupled with their nonplanar molecular structures, these new molecules should have considerable potential as materials for OLEDs (*14*–*17*). For example, as reported by Adachi, triphenylene-based molecules with bipyridine substituents are high–molecular orientation electron-transport materials (*15*, *16*) that have become one of the standard materials for OLEDs (*17*).

In addition to their potential use as optoelectronic materials, the annulation products also serve as partially fused polyaromatics and may be useful soluble precursors for the synthesis of fully fused, graphene nanoribbon substructures via Scholl-type dehydrocyclizations (*23*) (Fig. 4A). As a proof-of-concept, we used the annulative dimerization-dehydrocyclization sequence for the synthesis of fully fused nanographenes (Fig. 4B). Triphenylene **2a** was treated with $FeCl_3$ in CH_2Cl_2 at 0°C (standard conditions for the Scholl reaction), and **3** was obtained in 77% yield (62% two-step yield from **1a**). The clean, high-yielding formation of **3** without any optimization was quite surprising. It is well known that unwanted side reactions such as arene rearrangement (*24*) and aromatic chlorination often take place with polyaromatic compounds (*23*) in Scholl chemistry. For example, when polyphenylene **4**, which can be prepared by the nickel-catalyzed reductive dimerization of **1a** (*25*), was subjected to the standard Scholl conditions ($FeCl_3$, CH_2Cl_2, 0°C), the fully fused product **3** was not obtained; instead, a complex mixture of various unidentified products (most likely arene rearrangement products) was obtained. Thus, there is a clear advantage of the present reaction, accessing partially fused structures (triphenylene substructures in this case), to ensure that subsequent dehydrocyclization occurs successfully. This ring-fusing sequence, enabled by the annulative dimerization reaction, facilitates the synthesis of privileged fused π-conjugated systems, which are otherwise difficult to synthesize.

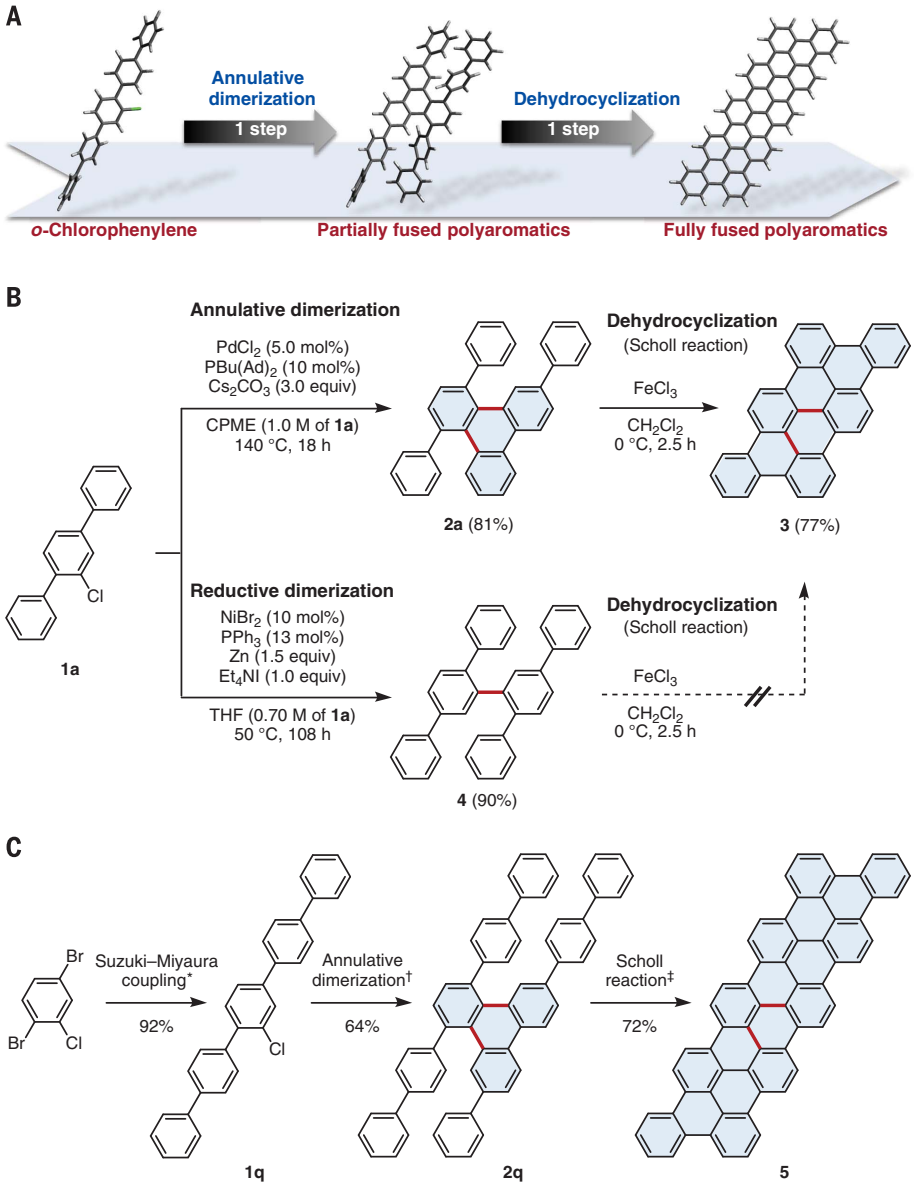


Fig. 4. Access to graphene nanoribbon substructures. (**A**) Rapid synthesis of partially and fully fused polyaromatics enabled by annulative dimerization of chlorophenylenes. (**B**) Advantage of annulative dimerization over classical reductive dimerization in the synthesis of fully fused phenylene **3**. THF, tetrahydrofuran; Et, ethyl. (**C**) Synthesis of graphene nanoribbon substructure **5**. Reaction conditions: *1,4-dibromo-2-chlorobenzene (1.0 equivalent), 4-biphenylboronic acid (2.4 equivalents), $PdCl_2(PPh_3)_2$ (2.0 mol %), K_2CO_3 (5.0 equivalents), toluene/H_2O/ethanol, 80°C, 18 hours, 92%. †$PdCl_2$ (5.0 mol %), $PBu(Ad)_2$ (10 mol %), Cs_2CO_3 (3.0 equivalents), CPME (1.0 M of **1q**), 140°C, 18 hours, 64%. ‡**2q** (0.010 mmol), $FeCl_3$ (0.47 mmol), CH_2Cl_2 (2.0 ml), room temperature, 42 hours, 72%.

To further demonstrate the utility of the present strategy, the rapid and convergent synthesis of an armchair-edged graphene nanoribbon segment (*3*, *26*–*31*) was carried out (Fig. 4C). 1,4-Dibromo-2-chlorobenzene was treated with *p*-biphenylylboronic acid in the presence of a palladium catalyst (Suzuki-Miyaura cross-coupling) to give chloropentaphenyl **1q** (92% yield). The palladium-catalyzed annulative dimerization of **1q** afforded partially fused product **2q** (64% yield). Finally, the Scholl reaction of **2q** gave small graphene nanoribbon segment $C_{60}H_{26}$ (**5**) in 72% yield. Notably, this 60-carbon nanoribbon **5** was obtained from 1,4-dibromo-2-chlorobenzene in three steps in 42% overall yield. The Raman and Fourier transform infrared spectra of small nanoribbon **5** were quite similar to the calculated spectra, supporting the formation of the expected nanoribbon structure (see figs. S7, S12, and S13 for details). This represents a very rare example of access to fully fused planar nanographenes without any solubilizing substituents under solution-phase conditions (*32*–*34*). Overall, the annulative dimerization reaction sequence reported here should drastically alter the execution of partially and fully fused polyaromatics, as chlorine and ortho hydrogen atoms on aromatic nuclei can now be considered ring-fusing handles.

REFERENCES AND NOTES

1. J. Hassan, M. Sévignon, C. Gozzi, E. Schulz, M. Lemaire, *Chem. Rev.* **102**, 1359–1470 (2002).
2. F. Ullmann, J. Bielecki, *Chem. Ber.* **34**, 2174–2185 (1901).
3. L. Chen, Y. Hernandez, X. Feng, K. Müllen, *Angew. Chem. Int. Ed.* **51**, 7640–7654 (2012).
4. L. T. Scott *et al.*, *J. Am. Chem. Soc.* **134**, 107–110 (2012).
5. G. Povie, Y. Segawa, T. Nishihara, Y. Miyauchi, K. Itami, *Science* **356**, 172–175 (2017).
6. M. G. Schwab *et al.*, *J. Am. Chem. Soc.* **134**, 18169–18172 (2012).
7. Y. Segawa, H. Ito, K. Itami, *Nat. Rev. Mater.* **1**, 15002 (2016).
8. L. Ackermann, R. Vicente, A. R. Kapdi, *Angew. Chem. Int. Ed.* **48**, 9792–9826 (2009).
9. Y. Segawa, T. Maekawa, K. Itami, *Angew. Chem. Int. Ed.* **54**, 66–81 (2015).
10. H. Ito, K. Ozaki, K. Itami, *Angew. Chem. Int. Ed.* **56**, 11144–11164 (2017).
11. K. Ozaki, K. Kawasumi, M. Shibata, H. Ito, K. Itami, *Nat. Commun.* **6**, 6251 (2015).
12. C.-W. Li, C.-I. Wang, H.-Y. Liao, R. Chaudhuri, R.-S. Liu, *J. Org. Chem.* **72**, 9203–9207 (2007).
13. R. K. Mohamed *et al.*, *Angew. Chem. Int. Ed.* **55**, 12054–12058 (2016).
14. P.-Y. Chou *et al.*, *Chem. Commun. (Camb.)* **50**, 6869–6871 (2014).
15. K. Togashi, S. Nomura, N. Yokoyama, T. Yasuda, C. Adachi, *J. Mater. Chem.* **22**, 20689–20695 (2012).
16. K. Togashi, Y. Sagara, T. Yasuda, C. Adachi, *Chem. Lett.* **42**, 651–653 (2013).
17. C. Adachi, K. Togashi, N. Yokoyama, S. Nomura, Patent WO 2012/114745 (2012).
18. S. L. Buchwald, B. T. Watson, J. C. Huffman, *J. Am. Chem. Soc.* **108**, 7411–7414 (1986).
19. D. A. Watson *et al.*, *Science* **325**, 1661–1664 (2009).
20. H. Jiang, Y. Zhang, D. Chen, B. Zhou, Y. Zhang, *Org. Lett.* **18**, 2032–2035 (2016).
21. M. A. Campo *et al.*, *J. Am. Chem. Soc.* **129**, 6298–6307 (2007).
22. L.-M. Xu, B.-J. Li, Z. Yang, Z.-J. Shi, *Chem. Soc. Rev.* **39**, 712–733 (2010).
23. M. Grzybowski, K. Skonieczny, H. Butenschön, D. T. Gryko, *Angew. Chem. Int. Ed.* **52**, 9900–9930 (2013).
24. J. He *et al.*, *Chem. Commun. (Camb.)* **51**, 7245–7248 (2015).
25. M. F. Semmelhack, P. M. Helquist, L. D. Jones, *J. Am. Chem. Soc.* **93**, 5908–5910 (1971).
26. K. Müllen, *ACS Nano* **8**, 6531–6541 (2014).
27. A. Narita *et al.*, *Nat. Chem.* **6**, 126–132 (2014).
28. L. T. Scott *et al.*, *Pure Appl. Chem.* **71**, 209–219 (1999).
29. V. M. Tsefrikas, L. T. Scott, *Chem. Rev.* **106**, 4868–4884 (2006).
30. J. Cai *et al.*, *Nature* **466**, 470–473 (2010).
31. M. Treier *et al.*, *Nat. Chem.* **3**, 61–67 (2011).
32. C. D. Simpson *et al.*, *Chem. Eur. J.* **8**, 1424–1429 (2002).
33. B. Schuler *et al.*, *Angew. Chem. Int. Ed.* **53**, 9004–9006 (2014).
34. Z. Wang *et al.*, *J. Am. Chem. Soc.* **126**, 7794–7795 (2004).

ACKNOWLEDGMENTS

This work was supported by JST ERATO grant number JPMJER1302 (K.I.) and JSPS KAKENHI grant numbers JP15K17821 and JP17H04868 (K.M.). We are grateful to Y. Segawa, H. Ito, and T. Yoshidomi for assistance with x-ray crystal structure analysis and fruitful discussions. We acknowledge Y. Miyauchi and A. Takakura for assistance with Raman measurements, H. Sakamoto and N. Ozaki for assistance with Fourier transform infrared measurements, and K. Kuwata and K. Ito for assistance with high-resolution mass spectrometry measurements. ITbM is supported by the World Premier International Research Center Initiative (WPI). Crystallographic data for compounds **2a**, **2p**, and **2p'** are available free of charge from the Cambridge Crystallographic Data Centre under CCDC identifiers 1518668, 1518669, and 1518670, respectively (www.ccdc.cam.ac.uk/structures/). All authors are inventors on a patent application (PCT/JP2017/020703) submitted by Nagoya University that covers the synthetic methods and molecules included in this paper.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6374/435/suppl/DC1
Materials and Methods
Figs. S1 to S188
Table S1
References (*35*–*47*)

16 September 2017; accepted 13 December 2017
10.1126/science.aap9801

Cite as: D. M. Zajac *et al.*, *Science* 10.1126/science.aao5965 (2017).

Resonantly driven CNOT gate for electron spins

D. M. Zajac,[1] A. J. Sigillito,[1] M. Russ,[2] F. Borjans,[1] J. M. Taylor,[3,4] G. Burkard,[2] J. R. Petta[1]*

[1]Department of Physics, Princeton University, Princeton, NJ 08544, USA. [2]Department of Physics, University of Konstanz, D-78457 Konstanz, Germany. [3]Joint Quantum Institute and Joint Center for Quantum Information and Computer Science, NIST and University of Maryland, College Park, MD 20742, USA. [4]Research Center for Advanced Science and Technology, University of Tokyo, Tokyo, Japan.

***Corresponding author. Email: petta@princeton.edu**

Single qubit rotations and two-qubit CNOT operations are crucial ingredients for universal quantum computing. Although high fidelity single qubit operations have been achieved using the electron spin degree of freedom, realizing a robust CNOT gate has been challenging owing to rapid nuclear spin dephasing and charge noise. We demonstrate an efficient resonantly driven CNOT gate for electron spins in silicon. Our platform achieves single-qubit rotations with fidelities >99%, as verified by randomized benchmarking. Gate control of the exchange coupling allows a quantum CNOT gate to be implemented with resonant driving in ~200 ns. We use the CNOT gate to generate a Bell state with 78% fidelity (corrected for errors in state preparation and readout). Our quantum dot device architecture enables multi-qubit algorithms in silicon.

Gate defined semiconductor quantum dots are a powerful platform for isolating and coherently controlling single electron spins (*1*, *2*). Silicon quantum dots can take advantage of state-of-the-art industrial nanofabrication capabilities for scalability, and support some of the longest quantum coherence times measured in the solid-state (*3*–*5*). By engineering local magnetic field gradients, electron spins can be electrically controlled (*6*, *7*) with single qubit gate fidelities exceeding 99.9% (*8*, *9*). Although exchange control of spins was demonstrated as early as 2005 (*10*), demonstrations of two-qubit gates with quantum dot spins are scarce owing to technological and materials challenges (*11*–*14*). A demonstration of an efficient CNOT gate for spins in silicon will open a path for multi-qubit algorithms in a scalable semiconductor system.

Here we demonstrate a ~200 ns CNOT gate in a silicon semiconductor double quantum dot (DQD), nearly an order of magnitude faster than the previously demonstrated composite CNOT gate (*14*). Turning on an exchange interaction results in a state-selective electron spin resonance (ESR) transition that is used to implement a CNOT gate with a single microwave (MW) pulse. In contrast with previous DQD implementations of the exchange gate, our CNOT gate is implemented at a symmetric operating point, where the exchange coupling J is first-order insensitive to charge noise (*15*–*17*).

We use the spin of a single electron to encode a qubit (*18*). A gate-defined DQD (Fig. 1A) isolates two electrons in a silicon quantum well with natural isotopic abundance, forming a two-qubit device (Fig. 1B). Gate L (R) is used to control the energy of the electron trapped in the left (right) quantum dot, and gate M provides control of J. The charge occupancy of the DQD is detected by monitoring the current I_S or conductance g_S through a nearby quantum dot charge sensor (*19*). A Co micromagnet (*20*) generates a magnetic field gradient that results in distinct ESR transition frequencies for the left and right qubits. Electrically driven single qubit rotations are implemented by "shaking" the electron spins in the transverse field gradient of the micromagnet (*7*, *21*). Other Si/SiGe device designs have suffered from accidental quantum dot formation, which is in part caused by the high effective mass of electrons in Si and gate designs that are much larger than the single electron wavefunction (*22*). Isolation and control of single electrons in our experiment is made possible by an overlapping aluminum gate electrode architecture that provides tight electronic confinement in a largely disorder-free potential. As we now demonstrate, these devices allow exquisite control of single electrons and nearest neighbor exchange coupling, with the potential to scale to at least 9 dots in a linear array (*19*).

The DQD gate voltages V_L and V_R are tuned to traverse the charge stability diagram. Starting from the charge state (N_L = 0, N_R = 0), where N_L (N_R) refers to the number of electrons in the left (right) dot, we navigate from points a to c in the stability diagram (Fig. 1C) to initialize the device in the $|\downarrow_L\downarrow_R\rangle$ state. The gates are then pulsed to point d, in the (1, 1) charge state, where single qubit control is achieved by applying MW pulses to gate S. Exchange can be rapidly turned on and off (on a 5 ns timescale) by adjusting the voltage V_M. Qubit readout is achieved by moving from points e to g, where the left and right dot spins are sequentially measured and emptied. Spin dependent tunneling and charge state readout are used to extract the spin-up probability $P_\uparrow^L$ ($P_\uparrow^R$) of the left (right) qubits following (*23*, *24*). Energy level diagrams corresponding to each point in the pulse sequence

are shown in Fig. 1D. High fidelity single qubit control is demonstrated in Fig. 2, with randomized benchmarking yielding single qubit fidelities F_L = 99.3 ± 0.2% and F_R = 99.7 ± 0.1%.

Proposals for two-qubit interactions with spins in semiconductors are generally based on control of the exchange coupling (*18*). In order to implement a high-fidelity CNOT gate we must first measure J as a function of V_M (*20*). Physically, in the presence of a strong magnetic field gradient $\delta B >> J$, the exchange interaction lowers the energy of the antiparallel spin states relative to the $|\uparrow\uparrow\rangle$ and $|\downarrow\downarrow\rangle$ spin states (Fig. 3A). As a result, the ESR frequency of the left qubit will be dependent on the state of the right qubit (and vice versa). We can therefore determine J by measuring the left qubit ESR spectra for different right qubit states (Fig. 3B). Specifically, the system is prepared in $|\downarrow\downarrow\rangle$ and then a rotation of duration τ_R is applied to the right qubit. Next we apply a low power probe tone for a time $\tau_L >> T_2$ at a frequency f_p that will leave the qubit in a mixed state if f_p is resonant with the qubit frequency. For the simple case where τ_R is such that the right qubit ends in the spin-down (spin-up) state, the left qubit will have a transition frequency $f^{L}_{|\psi_R\rangle=|\downarrow\rangle}$ $\left(f^{L}_{|\psi_R\rangle=|\uparrow\rangle}\right)$ as illustrated in the green (blue) box in Fig. 3B. By plotting $P^{L}_{\uparrow}$ as a function of τ_R and f_p (Fig. 3C) we see that the left qubit resonance frequency is correlated with the state of the right qubit (Fig. 3D). The exchange frequency $J/h = f^{L}_{|\psi_R\rangle=|\uparrow\rangle} - f^{L}_{|\psi_R\rangle=|\downarrow\rangle}$, where h is Planck's constant, is directly extracted from the data sets in Fig. 3E and plotted as a function of V_M in Fig. 3F (*20*). A 20 mV change in V_M is sufficient to turn on a 10 MHz exchange splitting, which exceeds typical single qubit Rabi frequencies (f_{Rabi} = 4.8 MHz in Fig. 2C).

Fast gate voltage control of J can be used to implement a resonant CNOT gate (Fig. 4). The general quantum circuit, and its experimental implementation, are shown in Fig. 4, A and B. When V_M is low, J is approximately zero (J ~ 300 kHz for V_M = 390 mV, see Fig. 3F) corresponding to the level diagram on the left in Fig. 4C (*20*). With J ~ 0 high fidelity single qubit gates can be implemented, because the resonance frequency of each qubit is independent of the state of the other qubit (see Fig. 2). When V_M is pulsed high, the antiparallel spin states are lowered in energy by $J/2$ relative to the parallel spin states (right panel, Fig. 4C). In this configuration, the CNOT gate is resonantly implemented by driving the $|\downarrow\uparrow\rangle$ to $|\uparrow\uparrow\rangle$ transition. Here $|tc\rangle$ describes a product state of the target (t) and control (c) qubits. In contrast, the $|\uparrow\downarrow\rangle$ to $|\downarrow\downarrow\rangle$ transition is off-resonant owing to the J coupling, allowing for a conditional rotation that is dependent on the right electron spin state.

To calibrate the CNOT gate, we use a long dc exchange pulse of length τ_{dc} = 1 μs and vary the length τ_P of the MW pulse to drive transitions between $|\downarrow\uparrow\rangle$ and $|\uparrow\uparrow\rangle$. The resulting conditional oscillations are shown in Fig. 4D for the input states $|\downarrow\uparrow\rangle$ and $|\downarrow\downarrow\rangle$. A conditional π-rotation is realized on the target qubit for $t_{CNOT} = \tau_P$ = 130 ns. Thanks to the magnetic field gradient, changes in V_M shift the orbital positions of the electrons and result in small changes in the ESR resonance frequencies. By setting $\tau_{dc} = 2\pi/J$ = 204 ns we eliminate conditional phases caused by exchange and the remaining single qubit phases are accounted for in the phase of the consecutive MW drives, resulting in a pure CNOT gate (*20*). In contrast to our single-step CNOT gate, implementation of a conventional CNOT gate following the Loss-DiVincenzo proposal would require mastery of two $\sqrt{SWAP}$ operations and three single qubit gates (*18*).

In general, the CNOT gate must be able to operate on an arbitrary input state, and specifically on product states of the form

$$|\psi_{in}\rangle = (\alpha_L|\downarrow\rangle_L + \beta_L|\uparrow\rangle_L) \otimes (\alpha_R|\downarrow\rangle_R + \beta_R|\uparrow\rangle_R) \quad (1)$$

where R and L denote the right (control) and left (target) qubits. Here $|\alpha_{L,R}|^2 + |\beta_{L,R}|^2 = 1$. To test the basic functionality of the CNOT gate, we first initialize the system in $|\downarrow\downarrow\rangle$. The control qubit is then rotated by an angle θ_R to create the input state

$$|\psi_{in}\rangle = |\downarrow\rangle_L \otimes \left[\cos\left(\frac{\theta_R}{2}\right)|\downarrow\rangle_R - i\sin\left(\frac{\theta_R}{2}\right)|\uparrow\rangle_R\right] \quad (2)$$

Figure 4E shows $P^{L}_{\uparrow}$ and $P^{R}_{\uparrow}$ measured after the CNOT gate acts on different input states with angle θ_R. These data show that the CNOT acts as expected on states outside of the classical set of (product) input states $|\uparrow\uparrow\rangle$, $|\uparrow\downarrow\rangle$, $|\downarrow\uparrow\rangle$, and $|\downarrow\downarrow\rangle$ and that the control qubit is not flipped during the CNOT operation.

We next use the CNOT gate (*20*) to create the Bell state $|\psi_{target}\rangle = \frac{1}{\sqrt{2}}(|\downarrow\downarrow\rangle - i|\uparrow\uparrow\rangle)$ and extract its fidelity by performing two qubit state tomography (*25*, *26*). The Bell state is created by initializing the system in $|\downarrow\downarrow\rangle$ and applying a $\pi/2_x$-pulse on the control spin, which generates the input state $|\psi_{in}\rangle = \frac{1}{\sqrt{2}}(|\downarrow\downarrow\rangle - i|\downarrow\uparrow\rangle)$. Application of the CNOT gate to $|\psi_{in}\rangle$ yields $|\psi_{target}\rangle = \frac{1}{\sqrt{2}}(|\downarrow\downarrow\rangle - i|\uparrow\uparrow\rangle)$. State tomography is performed by appending single qubit rotations after the CNOT gate to measure the expectation value for all two qubit Pauli operators (for example, by applying a $\pi/2_x$ rotation to the left qubit and $\pi/2_y$ rotation to the right we measure the YX two qubit operator). Because the set of Pauli

operators form a basis of the Hermitian operators on the two qubit Hilbert space, we can reconstruct the full two qubit density matrix from these measurements. Not accounting for imperfections in state preparation and measurement (SPAM), we obtain a fidelity $F_{raw} = \langle\psi_{target}|\rho|\psi_{target}\rangle = 56\%$ (*27*). The readout visibilities of both qubits (*20*), and spin relaxation during the sequential qubit readout, account for a significant amount of the infidelity. By adopting the procedure used in (*22*), we correct our density matrix to account for SPAM errors and obtain a corrected fidelity $F = \langle\psi_{target}|\rho|\psi_{target}\rangle = 78\%$. We anticipate that the state fidelity can be significantly improved through the use of optimized pulse sequences (*28*).

Realizing robust two-qubit gates has been a bottleneck in the development of spin-based quantum computers (*12*, *29*). By combining our results with recent advances in Si/SiGe quantum dot device technology (*19*), we anticipate that it will now be feasible to demonstrate 9 qubit quantum processors in silicon.

REFERENCES AND NOTES

1. R. Hanson, L. P. Kouwenhoven, J. R. Petta, S. Tarucha, L. M. K. Vandersypen, Spins in few-electron quantum dots. *Rev. Mod. Phys.* **79**, 1217–1265 (2007). doi:10.1103/RevModPhys.79.1217

2. F. A. Zwanenburg, A. S. Dzurak, A. Morello, M. Y. Simmons, L. C. L. Hollenberg, G. Klimeck, S. Rogge, S. N. Coppersmith, M. A. Eriksson, Silicon quantum electronics. *Rev. Mod. Phys.* **85**, 961–1019 (2013). doi:10.1103/RevModPhys.85.961

3. M. Steger, K. Saeedi, M. L. W. Thewalt, J. J. L. Morton, H. Riemann, N. V. Abrosimov, P. Becker, H.-J. Pohl, Quantum information storage for over 180 s using donor spins in a ^{28}Si "semiconductor vacuum". *Science* **336**, 1280–1283 (2012). doi:10.1126/science.1217635 Medline

4. A. M. Tyryshkin, S. Tojo, J. J. L. Morton, H. Riemann, N. V. Abrosimov, P. Becker, H.-J. Pohl, T. Schenkel, M. L. W. Thewalt, K. M. Itoh, S. A. Lyon, Electron spin coherence exceeding seconds in high-purity silicon. *Nat. Mater.* **11**, 143–147 (2011). doi:10.1038/nmat3182 Medline

5. M. Veldhorst, J. C. C. Hwang, C. H. Yang, A. W. Leenstra, B. de Ronde, J. P. Dehollain, J. T. Muhonen, F. E. Hudson, K. M. Itoh, A. Morello, A. S. Dzurak, An addressable quantum dot qubit with fault-tolerant control-fidelity. *Nat. Nanotechnol.* **9**, 981–985 (2014). doi:10.1038/nnano.2014.216 Medline

6. M. Pioro-Ladrière, T. Obata, Y. Tokura, Y.-S. Shin, T. Kubo, K. Yoshida, T. Taniyama, S. Tarucha, Electrically driven single-electron spin resonance in a slanting Zeeman field. *Nat. Phys.* **4**, 776–779 (2008). doi:10.1038/nphys1053

7. Y. Tokura, W. G. van der Wiel, T. Obata, S. Tarucha, Coherent single electron spin control in a slanting Zeeman field. *Phys. Rev. Lett.* **96**, 047202 (2006). doi:10.1103/PhysRevLett.96.047202 Medline

8. J. P. Dehollain, J. T. Muhonen, R. Blume-Kohout, K. M. Rudinger, J. K. Gamble, E. Nielsen, A. Laucht, S. Simmons, R. Kalra, A. S. Dzurak, A. Morello, Optimization of a solid-state electron spin qubit using gate set tomography. *New J. Phys.* **18**, 103018 (2016). doi:10.1088/1367-2630/18/10/103018

9. J. Yoneda, K. Takeda, T. Otsuka, T. Nakajima, M. R. Delbecq, G. Allison, T. Honda, T. Kodera, S. Oda, Y. Hoshi, N. Usami, K. M. Itoh, S. Tarucha, A >99.9%-fidelity quantum-dot spin qubit with coherence limited by charge noise. arXiv:1708.01454 (2017).

10. J. R. Petta, A. C. Johnson, J. M. Taylor, E. A. Laird, A. Yacoby, M. D. Lukin, C. M. Marcus, M. P. Hanson, A. C. Gossard, Coherent manipulation of coupled electron spins in semiconductor quantum dots. *Science* **309**, 2180–2184 (2005). doi:10.1126/science.1116955 Medline

11. K. C. Nowack, M. Shafiei, M. Laforest, G. E. D. K. Prawiroatmodjo, L. R. Schreiber, C. Reichl, W. Wegscheider, L. M. K. Vandersypen, Single-shot correlations and two-qubit gate of solid-state spins. *Science* **333**, 1269–1272 (2011). doi:10.1126/science.1209524 Medline

12. R. Brunner, Y.-S. Shin, T. Obata, M. Pioro-Ladrière, T. Kubo, K. Yoshida, T. Taniyama, Y. Tokura, S. Tarucha, Two-qubit gate of combined single-spin rotation and interdot spin exchange in a double quantum dot. *Phys. Rev. Lett.* **107**, 146801 (2011). doi:10.1103/PhysRevLett.107.146801 Medline

13. M. D. Shulman, O. E. Dial, S. P. Harvey, H. Bluhm, V. Umansky, A. Yacoby, Demonstration of entanglement of electrostatically coupled singlet-triplet qubits. *Science* **336**, 202–205 (2012). doi:10.1126/science.1217692 Medline

14. M. Veldhorst, C. H. Yang, J. C. C. Hwang, W. Huang, J. P. Dehollain, J. T. Muhonen, S. Simmons, A. Laucht, F. E. Hudson, K. M. Itoh, A. Morello, A. S. Dzurak, A two-qubit logic gate in silicon. *Nature* **526**, 410–414 (2015). doi:10.1038/nature15263 Medline

15. F. Martins, F. K. Malinowski, P. D. Nissen, E. Barnes, S. Fallahi, G. C. Gardner, M. J. Manfra, C. M. Marcus, F. Kuemmeth, Noise suppression using symmetric exchange gates in spin qubits. *Phys. Rev. Lett.* **116**, 116801 (2016). doi:10.1103/PhysRevLett.116.116801 Medline

16. M. D. Reed, B. M. Maune, R. W. Andrews, M. G. Borselli, K. Eng, M. P. Jura, A. A. Kiselev, T. D. Ladd, S. T. Merkel, I. Milosavljevic, E. J. Pritchett, M. T. Rakher, R. S. Ross, A. E. Schmitz, A. Smith, J. A. Wright, M. F. Gyure, A. T. Hunter, Reduced sensitivity to charge noise in semiconductor spin qubits via symmetric operation. *Phys. Rev. Lett.* **116**, 110402 (2016). doi:10.1103/PhysRevLett.116.110402 Medline

17. B. Bertrand, H. Flentje, S. Takada, M. Yamamoto, S. Tarucha, A. Ludwig, A. D. Wieck, C. Bäuerle, T. Meunier, Quantum manipulation of two-electron spin states in isolated double quantum dots. *Phys. Rev. Lett.* **115**, 096801 (2015). doi:10.1103/PhysRevLett.115.096801 Medline

18. D. Loss, D. P. DiVincenzo, Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120–126 (1998). doi:10.1103/PhysRevA.57.120

19. D. M. Zajac, T. M. Hazard, X. Mi, E. Nielsen, J. R. Petta, Scalable gate architecture for a one-dimensional array of semiconductor spin qubits. *Phys. Rev. Appl.* **6**, 054013 (2016). doi:10.1103/PhysRevApplied.6.054013

20. See supplementary materials.

21. E. Kawakami, P. Scarlino, D. R. Ward, F. R. Braakman, D. E. Savage, M. G. Lagally, M. Friesen, S. N. Coppersmith, M. A. Eriksson, L. M. K. Vandersypen, Electrical control of a long-lived spin qubit in a Si/SiGe quantum dot. *Nat. Nanotechnol.* **9**, 666–670 (2014). doi:10.1038/nnano.2014.153 Medline

22. T. F. Watson, S. G. J. Philips, E. Kawakami, D. R. Ward, P. Scarlino, M. Veldhorst, D. E. Savage, M. G. Lagally, M. Friesen, S. N. Coppersmith, M. A. Eriksson, L. M. K. Vandersypen, A programmable two-qubit quantum processor in silicon. arXiv:1708.04214 (2017).

23. J. M. Elzerman, R. Hanson, L. H. Willems Van Beveren, B. Witkamp, L. M. K. Vandersypen, L. P. Kouwenhoven, Single-shot read-out of an individual electron spin in a quantum dot. *Nature* **430**, 431–435 (2004). doi:10.1038/nature02693 Medline

24. A. Morello, J. J. Pla, F. A. Zwanenburg, K. W. Chan, K. Y. Tan, H. Huebl, M. Möttönen, C. D. Nugroho, C. Yang, J. A. van Donkelaar, A. D. C. Alves, D. N. Jamieson, C. C. Escott, L. C. L. Hollenberg, R. G. Clark, A. S. Dzurak, Single-shot readout of an electron spin in silicon. *Nature* **467**, 687–691 (2010). doi:10.1038/nature09392 Medline

25. M. Steffen, M. Ansmann, R. C. Bialczak, N. Katz, E. Lucero, R. McDermott, M. Neeley, E. M. Weig, A. N. Cleland, J. M. Martinis, Measurement of the entanglement of two superconducting qubits via state tomography. *Science* **313**, 1423–1425 (2006). doi:10.1126/science.1130886 Medline

26. Y. X. Liu, L. F. Wei, F. Nori, Tomographic measurements on superconducting qubit states. *Phys. Rev. B* **72**, 014547 (2005). doi:10.1103/PhysRevB.72.014547

27. M. A. Nielsen, I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2000).

28. M. Russ, D. M. Zajac, A. J. Sigillito, F. Borjans, J. M. Taylor, J. R. Petta, G. Burkhard, High-fidelity quantum gates in Si/SiGe double quantum dots. arXiv:1711.00754 (2017).

29. Shortly after the submission of our results related work by another group appeared demonstrating exchange coupling in a Si/SiGe double quantum dot (*22*).

30. J. Yoneda, T. Otsuka, T. Takakura, M. Pioro-Ladrière, R. Brunner, H. Lu, T. Nakajima, T. Obata, A. Noiri, C. J. Palmstrøm, A. C. Gossard, S. Tarucha, Robust micromagnet design for fast electrical manipulations of single spins in quantum dots. *Appl. Phys. Express* **8**, 084401 (2015). doi:10.7567/APEX.8.084401

31. T. Meunier, V. E. Calado, L. M. K. Vandersypen, Efficient controlled-phase gate for single-spin qubits in quantum dots. *Phys. Rev. B* **83**, 121403 (2011). doi:10.1103/PhysRevB.83.121403

32. E. Magesan, J. M. Gambetta, J. Emerson, Scalable and robust randomized benchmarking of quantum processes. *Phys. Rev. Lett.* **106**, 180504 (2011). doi:10.1103/PhysRevLett.106.180504 Medline

33. J. T. Muhonen, A. Laucht, S. Simmons, J. P. Dehollain, R. Kalra, F. E. Hudson, S. Freer, K. M. Itoh, D. N. Jamieson, J. C. McCallum, A. S. Dzurak, A. Morello, Quantifying the quantum gate fidelity of single-atom spin qubits in silicon by randomized benchmarking. *J. Phys. Condens. Matter* **27**, 154205 (2015). doi:10.1088/0953-8984/27/15/154205 Medline

34. J. Kelly, R. Barends, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, I.-C. Hoi, E. Jeffrey, A. Megrant, J. Mutus, C. Neill, P. J. J. O'Malley, C. Quintana, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, A. N. Cleland, J. M. Martinis, Optimal quantum control using randomized benchmarking. *Phys. Rev. Lett.* **112**, 240504 (2014). doi:10.1103/PhysRevLett.112.240504 Medline

35. D. J. Reilly, C. M. Marcus, M. P. Hanson, A. C. Gossard, Fast single-charge sensing with a rf quantum point contact. *Appl. Phys. Lett.* **91**, 162101 (2007). doi:10.1063/1.2794995

36. M. J. Curry, T. D. England, N. C. Bishop, G. Ten-Eyck, J. R. Wendt, T. Pluym, M. P. Lilly, S. M. Carr, M. S. Carroll, Cryogenic preamplification of a single-electron-transistor using a silicon-germanium heterojunction-bipolar-transistor. *Appl. Phys. Lett.* **106**, 203505 (2015). doi:10.1063/1.4921308

37. J. Stehlik, Y.-Y. Liu, C. M. Quintana, C. Eichler, T. R. Hartke, J. R. Petta, Fast charge sensing of a cavity-coupled double quantum dot using a Josephson parametric amplifier. *Phys. Rev. Appl.* **4**, 014018 (2015). doi:10.1103/PhysRevApplied.4.014018

38. W. B. Mims, Measurement of the linear electric field effect in EPR using the spin echo method. *Rev. Sci. Instrum.* **45**, 1583 (1974). doi:10.1063/1.1686567

39. D. F. V. James, P. G. Kwiat, W. J. Munro, A. G. White, Measurement of qubits. *Phys. Rev. A* **64**, 052312 (2001). doi:10.1103/PhysRevA.64.052312

40. Z. Hradil, Quantum-state estimation. *Phys. Rev. A* **55**, R1561–R1564 (1997). doi:10.1103/PhysRevA.55.R1561

ACKNOWLEDGMENTS

We thank T. Hazard, J. Stehlik, and K. Wang for technical assistance. Research was sponsored by Army Research Office grant W911NF-15-1-0149, the Gordon and Betty Moore Foundation's EPiQS Initiative through grant GBMF4535, and NSF grant DMR-1409556. Devices were fabricated in the Princeton University Quantum Device Nanofabrication Laboratory. Any mention of commercial products is for information only; it does not imply recommendation or endorsement by NIST. The data that support the findings of this study are available in the supplementary material. Additional data can be obtained from the corresponding author upon request.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/cgi/content/full/science.aao5965/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S10
References (*30*–*40*)

7 August 2017; accepted 28 November 2017
Published online 7 December 2017
10.1126/science.aao5965

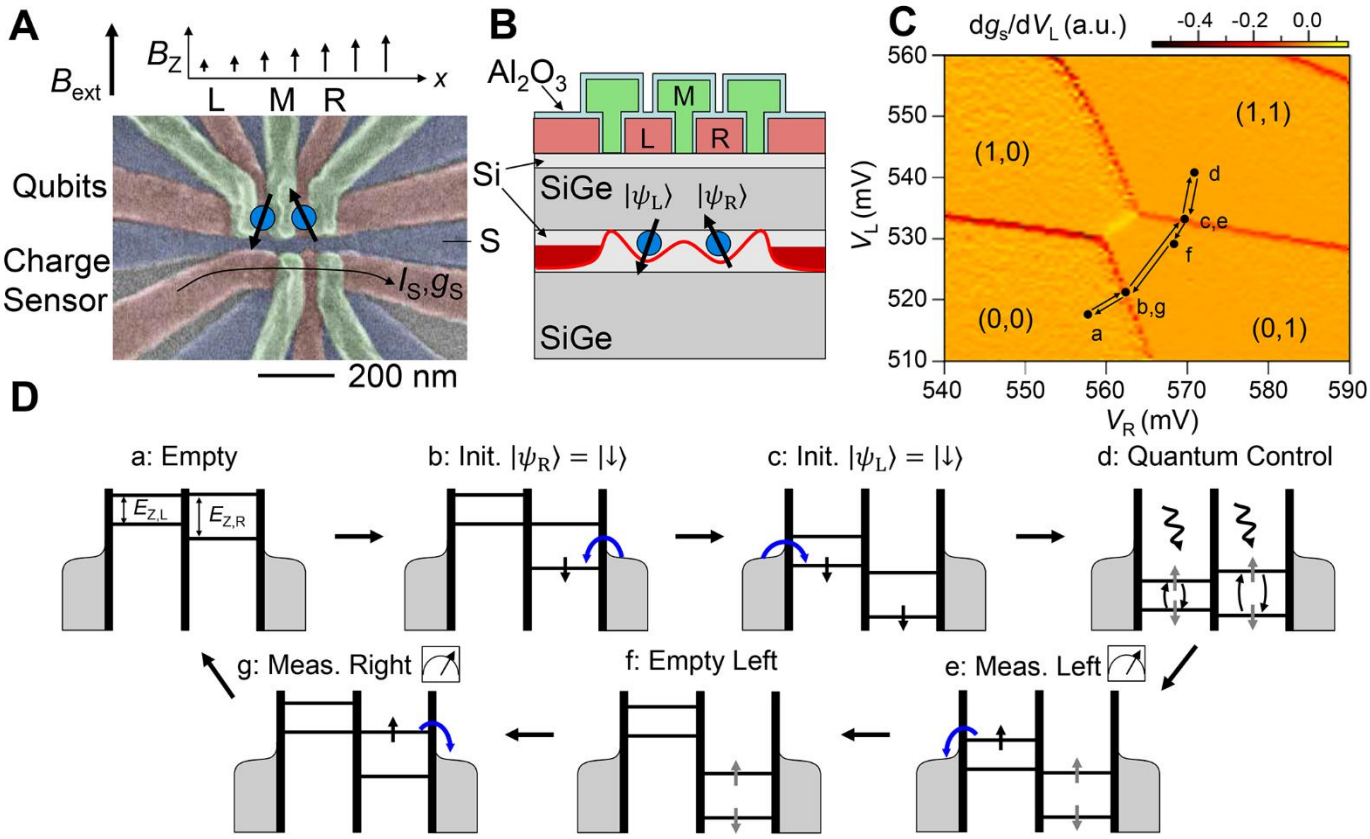


Fig. 1. Two-qubit device. (**A**) False-color scanning electron microscope image of the DQD before the deposition of the Co micromagnet. Two spin qubits are defined by the DQD and a neighboring quantum dot is used as a charge sensor. The Co micromagnet (not shown) creates a slanting Zeeman field that is used for quantum control (upper schematic). (**B**) Schematic cross-section of the DQD device. Two electrons are trapped in the confinement potential created by gates L, M, and R, to which voltages V_L, V_M, and V_R are applied, respectively. (**C**) DQD charge stability diagram. Points a-g are used in the two-qubit control sequence. (**D**) DQD energy level configuration at each point in the pulse sequence. Points a-c are used to initialize the system in $|\downarrow\downarrow\rangle$. Single qubit and two qubit gates are implemented at point d. Sequential single-shot spin state readout is achieved by navigating from points e-g.

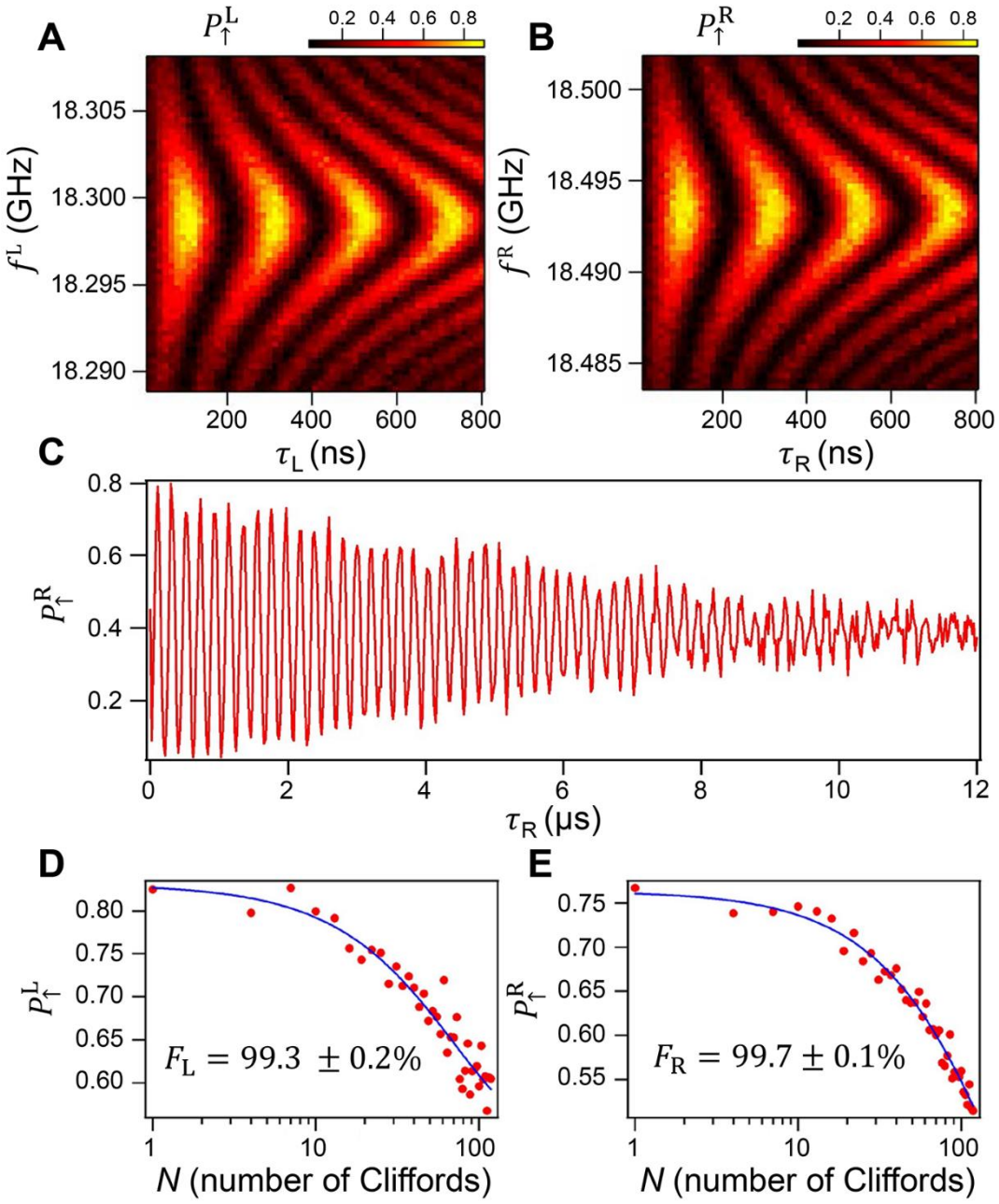


Fig. 2. High fidelity single qubit gates. (A) Left qubit spin-up probability $P_\uparrow^L$ plotted as a function of the MW drive frequency f^L and drive time τ_L, showing coherent Rabi oscillations. (B) Right qubit Rabi oscillations. (C) $P_\uparrow^R$ as a function of τ_R, obtained at a drive frequency f^R = 18.493 GHz, shows high visibility Rabi oscillations that persist to 10 μs. Clifford randomized benchmarking (*20*) of the left (D) and right (E) qubits yields gate fidelities F_L and F_R in excess of 99%.

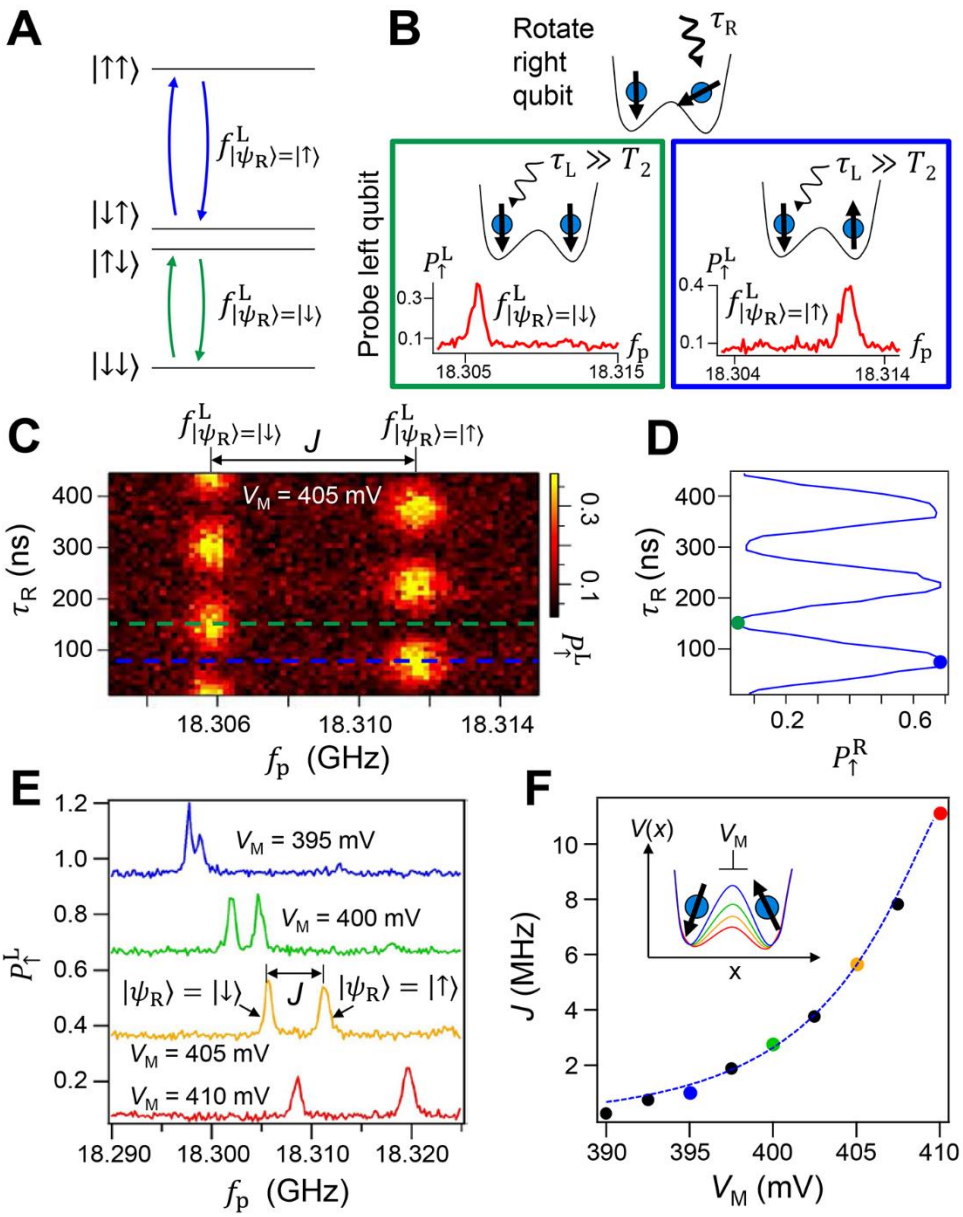


Fig. 3. Exchange spectroscopy. (A) Schematic energy level diagram with a large J. (B) We spectroscopically measure J by first applying a rotation to the right qubit and then applying a low power probe tone to the left qubit. The left qubit will have a resonance frequency $f^{L}_{|\psi_R\rangle=|\downarrow\rangle}$ when the right qubit is in the spin-down state (green box) and $f^{L}_{|\psi_R\rangle=|\uparrow\rangle}$ when the right qubit is in the spin-up state (blue box). (C) $P_\uparrow^L$ as a function of τ_R and the probe frequency f_p. The two resonance frequencies of the left qubit are split by J. The response of the left qubit to the probe tone oscillates between these two frequencies as the right qubit oscillates between spin-up and spin-down. (D) $P_\uparrow^R$ as a function of τ_R, displaying Rabi oscillations. (E) Spectra showing the left dot resonance frequencies for four different values of V_M. Curves offset by 0.3 for clarity. (F) J as a function of V_M (dots) and theory predictions (line) from (*20*).

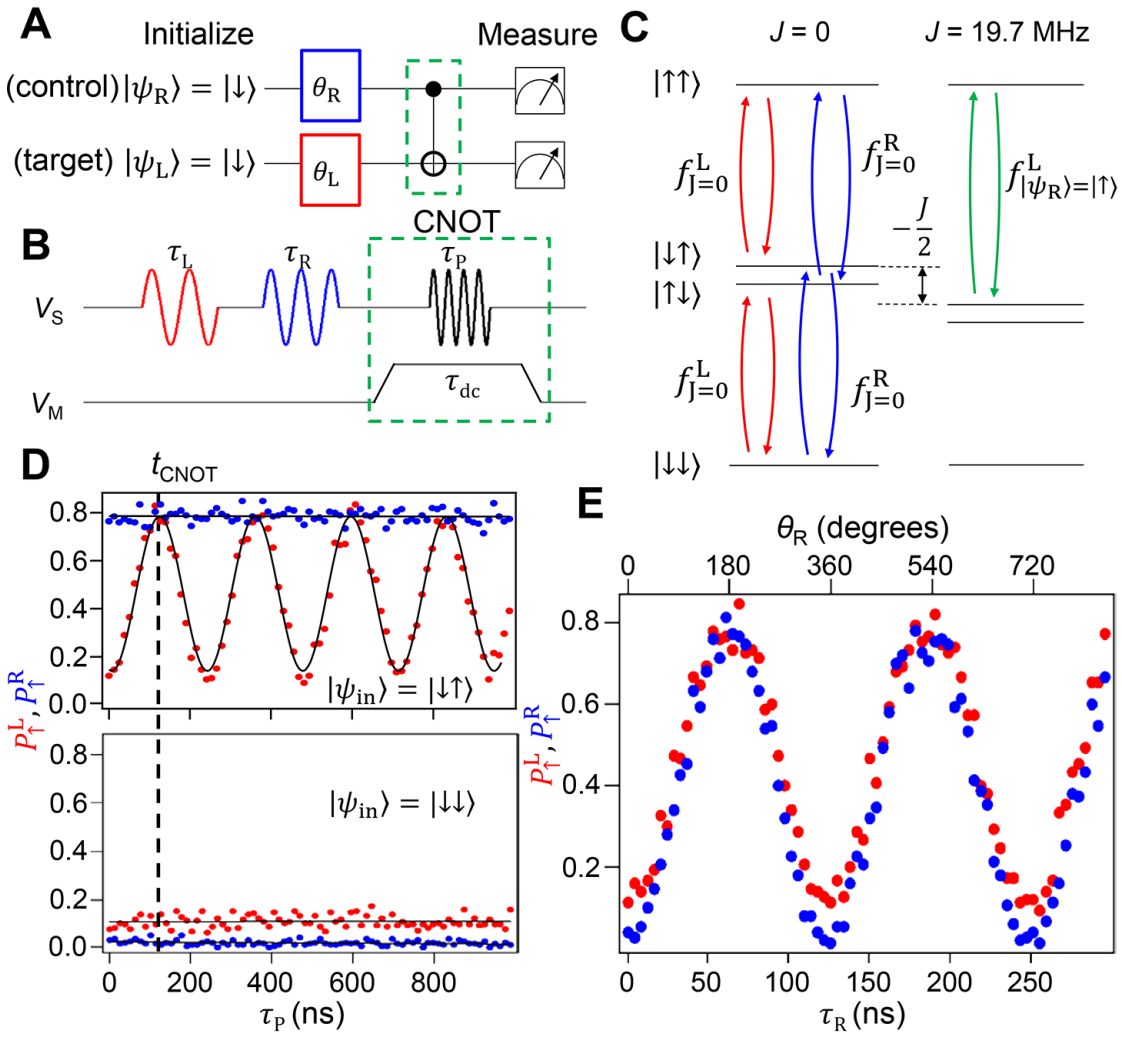


Fig. 4. CNOT gate. (A) Quantum circuit for the CNOT gate. (B) Experimental implementation of the quantum circuit. (C) Schematic energy level diagrams for $J = 0$ (left) and $J \neq 0$ (right). When $J \neq 0$ a conditional rotation can be applied to the left qubit by driving at $f^{L}_{|\psi_R\rangle=|\uparrow\rangle}$. (D) $P^{L}_{\uparrow}$ (red) and $P^{R}_{\uparrow}$ (blue) as a function of τ_P for input states $|\downarrow\uparrow\rangle$ (top) and $|\downarrow\downarrow\rangle$ (bottom). The vertical dashed line at $\tau_P = 130$ ns $= t_{CNOT}$ corresponds to a CNOT gate. (E) CNOT gate with superposition input states. With $\tau_P = 130$ ns and $\tau_L = 0$, the response of the CNOT gate is plotted as a function of τ_R showing that the target qubit follows the state of the control qubit.

NANOPHOTONICS

Nanoscale chiral valley-photon interface through optical spin-orbit coupling

Su-Hyun Gong,[1,2] Filippo Alpeggiani,[1,2] Beniamino Sciacca,[2] Erik C. Garnett,[2] L. Kuipers[1,2]*

The emergence of two-dimensional transition metal dichalcogenide materials has sparked intense activity in valleytronics, as their valley information can be encoded and detected with the spin angular momentum of light. We demonstrate the valley-dependent directional coupling of light using a plasmonic nanowire–tungsten disulfide (WS_2) layers system. We show that the valley pseudospin in WS_2 couples to transverse optical spin of the same handedness with a directional coupling efficiency of 90 ± 1%. Our results provide a platform for controlling, detecting, and processing valley and spin information with precise optical control at the nanoscale.

Valleytronics and nanophotonics provide powerful routes to address the heating problem in electronics by offering an alternative to information transport with the charge of electrons (*1*, *2*). Valley pseudospin provides an additional degree of freedom to encode and process binary information in matter, analogous to the spin degree of freedom. The emergence of two-dimensional transition metal dichalcogenide (TMD) layers provides a versatile materials platform for both optoelectronics (*3*) and valleytronics (*4*–*9*). These materials have direct band gaps consisting of two (energy-degenerate) valleys at the corners of the Brillouin zone (labeled K and K′). The spin and valley information in TMD materials can be optically addressed and detected by using the spin angular momentum of light because of their valley-dependent optical selection rule (*7*, *8*, *10*–*12*). However, the relatively short lifetimes (<10 ps) of valley-polarized excitons limit logical processes and the spatial transport of valley information (*2*). Although valley information is difficult to transport in realistic material systems, particularly at room temperature, photons are ideal information carriers owing to the normally small light-matter interaction. Nanoscale coupling of valley pseudospin to photonic degrees of freedom is desirable for on-chip integrated valley devices.

The spin angular momentum as a degree of freedom of light at the nanoscale offers the ability to influence and exploit light-matter interactions. In highly confined light fields, transverse optical spin angular momentum (t-OSAM) results from the spin-orbit interaction of light (*13*–*25*). This t-OSAM provides a robust one-to-one relation between the handedness of optical spin and the propagation direction of a photonic mode—so-called spin-momentum locking—because of their time reversal symmetry. The information of spin angular momentum can be directly transferred to the direction of light and vice versa. Recently, t-OSAM–dependent directional coupling of light with near-unity efficiency has been successfully demonstrated experimentally by using various photonic structures, e.g., optical fibers (*13*–*15*), metal surfaces (*16*–*18*), metasurfaces (*26*), semiconductor waveguides (*20*, *21*), and microdisks (*22*).

We demonstrate a room-temperature chiral-coupling interface between the transverse optical spin of a plasmonic nanowire mode and the valley pseudospin of tungsten disulfide (WS_2), which provides robust valley-polarized directional emission. The resulting coupling between the photonic path and valley-spin in TMD mateials and the metallic nature of the waveguide should enable applications in valley-chiral networks, valley-gates, and quantum photonic devices.

Our configuration for a chiral valley-photon interface (Fig. 1A) consists of a combination of a few-layer TMD material and a single silver nanowire. The evanescent fields of the plasmonic guided modes possess t-OSAM perpendicular to the TMD layers. The population of excitons in the two different valleys K and K′ can be directly controlled by a circularly polarized excitation laser. The radiative decay of a valley-polarized exciton is associated with a circular transition dipole, which emits valley-dependent circularly polarized light. This circular transition dipole couples to the plasmonic eigenstate with the same handedness of its local transverse optical spin. As a result of this chiral coupling between the circular nature of the emission dipole of the valley-polarized exciton and the local direction-locked transverse optical spin, emission from the different valleys couples to plasmonic modes propagating in opposite directions (Fig. 1B) when the transverse optical spin is maximal. To obtain a high degree of valley polarization, we exploit the high spin-valley coupling strength and spin-layer locking effect of WS_2 (*27*, *28*). The degree of valley polarization is denoted by $P_V = (I_K - I_{K'})/(I_K + I_{K'})$, where I_K and $I_{K'}$ represent the photoluminescence (PL) intensity from K and K′, respectively. The number of WS_2 layers is identified by the optical contrast of the layers in a microscope image and their PL spectrum (fig. S1). The results presented here are based on a WS_2 flake consisting of five layers. Its polarization-resolved spectrum, measured in the absence of the silver nanowire (Fig. 1C), reveals a measured $P_V = 0.7$ at room temperature. Figure 1D depicts an optical microscopy image of one of the samples and a collected PL image dominated by emission from the indirect band gap. The two bright spots at each end of the waveguide demonstrate that emission is coupled to the guided modes of the plasmonic nanowire. Through the chiral coupling, the valley information is converted to the propagation direction of the mode to which the exciton coupled: The plasmonic mode itself has no overall optical spin (fig. S2). Under local excitation at the middle of the silver nanowire, directional emission from the TMD layers is investigated by measuring light scattered at the ends of the wire.

To quantify the magnitude of the t-OSAM near a silver nanowire, we performed numerical calculations with a finite-difference eigenmode solver (Fig. 2A). The guided plasmonic mode exhibits strong evanescent fields at the glass-nanowire interface, i.e., exactly at the position of WS_2 layers. Due to the strong transverse confinement and the plasmonic nature of the modes, a large longitudinal (x direction) component of the electric field is present. The x and y components of the electric field have comparable amplitudes and are roughly ±90° out of phase: The modes exhibit a large t-OSAM. The local sign of the local t-OSAM has a one-to-one relation with the propagation direction of light and the position with respect to the mirror plane of the geometry (y direction). We calculated the density of the t-OSAM as a function of position in the x-y plane, which corresponds to the Stokes parameter $S_3 = -2\mathrm{Im}(E_x E_y^*)/(|E_x|^2 + |E_y|^2)$ (*20*) (Fig. 2B). The plasmonic guided modes of infinite-length nanowires locally have a t-OSAM density near unity. The sign of the transverse optical spin on either side of the nanowire is opposite, as expected from symmetry considerations. Clearly, when the propagation direction of the mode is reversed, so are all the signs of the helicity everywhere. These results are qualitatively summarized in Fig. 2C, demonstrating that the handedness of optical spin is determined by a combination of y position and propagation direction.

We simulate the emission from one valley in WS_2 near a finite-length plasmonic nanowire

[1]Kavli Institute of Nanoscience, Department of Quantum Nanoscience, Delft University of Technology, Post Office Box 5046, 2600 GA Delft, Netherlands. [2]Center for Nanophotonics, AMOLF, Science Park 104, 1098 XG Amsterdam, Netherlands.
*Corresponding author. Email: l.kuipers@tudelft.nl

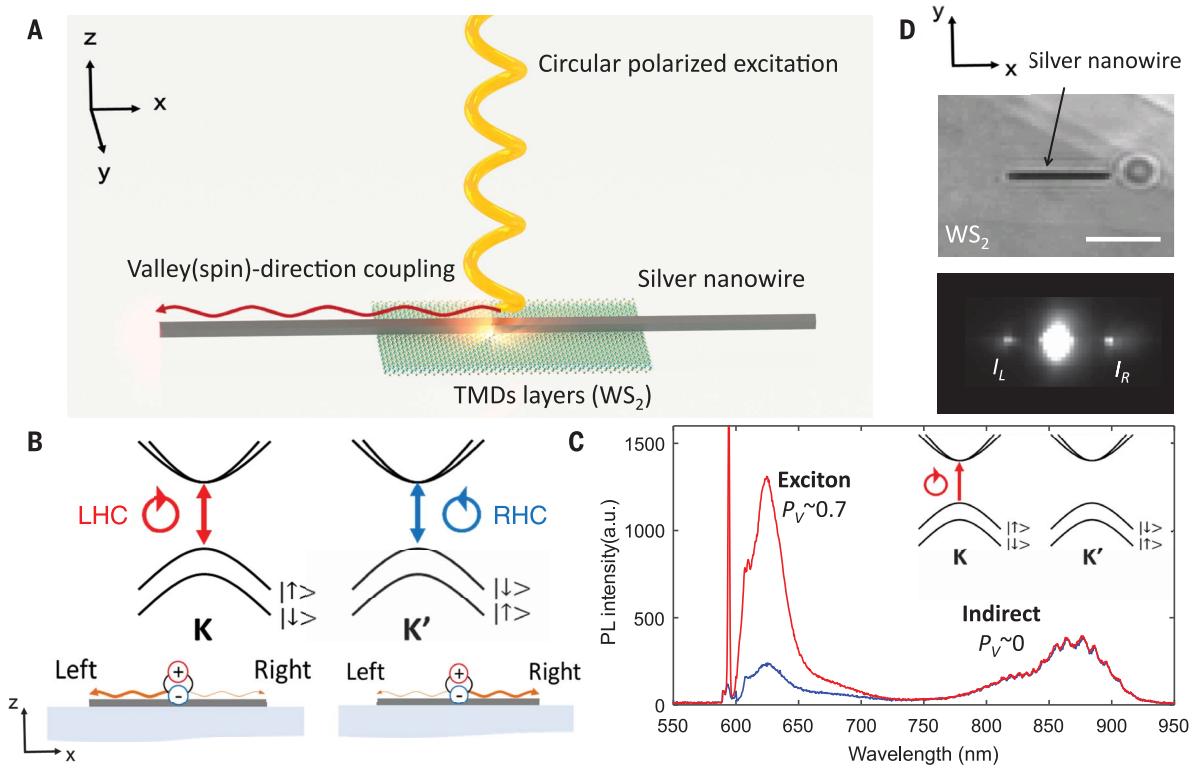


Fig. 1. Valley-controlled directional coupling of light. (**A**) A conceptual illustration of directional emission of a valley-polarized exciton in WS_2. The valley pseudospin and photon path are coupled by means of spin-orbit coupling of light. (**B**) A schematic sketch of the band diagram of WS_2 and its optical selection rules, which depend on the valley index. The two opposite-handed circularly polarized emissions from each valley would couple preferentially to modes propagating in opposite directions. LHC, left-handed circular; RHC, right-handed circular. (**C**) Polarization-resolved emission spectrum of WS_2, measured without a silver nanowire, at room temperature excited by a left-handed circularly polarized laser beam at 594 nm. Red and blue spectra indicate left- and right-handed circularly polarized emission, respectively. a.u., arbitrary units. (**D**) Image of a fabricated WS_2–silver nanowire coupled system (top). WS_2 emission coupled to the plasmonic waveguide mode and scattered at the ends of the nanowire (bottom). Scale bar, 5 μm.

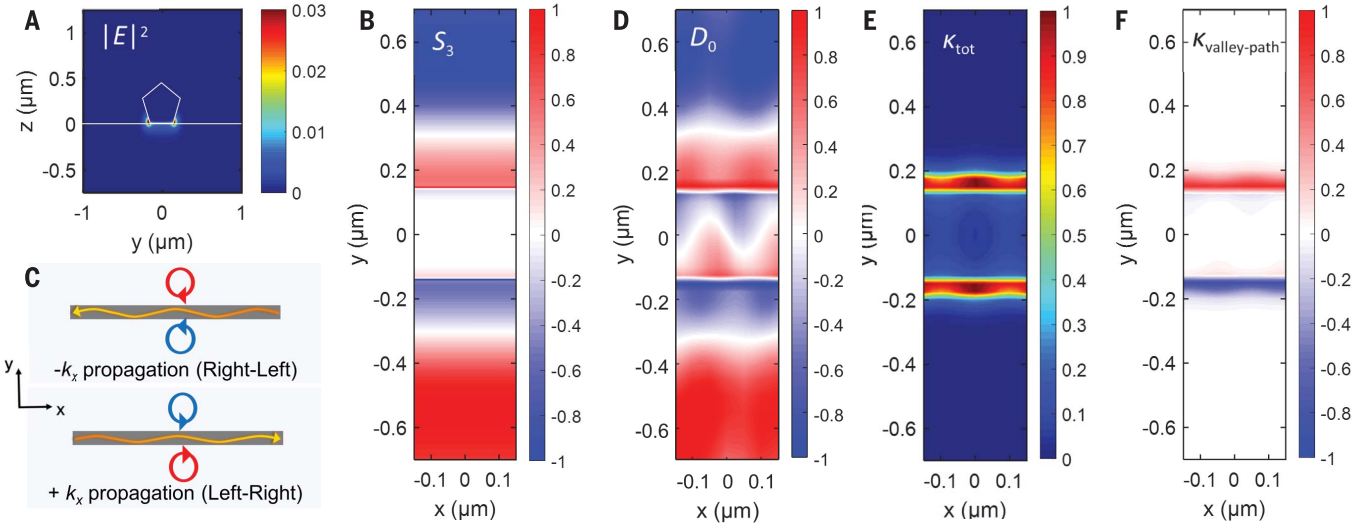


Fig. 2. Numerical modeling of the transverse spin angular momentum of light. (**A**) Cross-sectional electric field intensity ($|E|^2$) distribution of the plasmonic guided mode. (**B**) Density of transverse optical spin, i.e., Stokes parameter S_3, of the in-plane (x-y plane) electric field component. (**C**) Illustration of the distribution of handedness of elliptical polarization, optical spin, which is dependent on both the position and propagation direction. (**D**) Directionality of circularly polarized dipole emission, $D_0(x,y)$, as a function of its position. (**E**) Normalized coupling strength of circularly polarized dipole source to the plasmonic guided modes, $\kappa_{tot}(x,y)$. (**F**) Chiral-coupling coefficient indicating how efficient circular dipole in one valley couples to one direction of the guided mode, $\kappa_{valley\text{-}path}(x,y) = D_0(x,y)\cdot\kappa_{tot}(x,y)$.

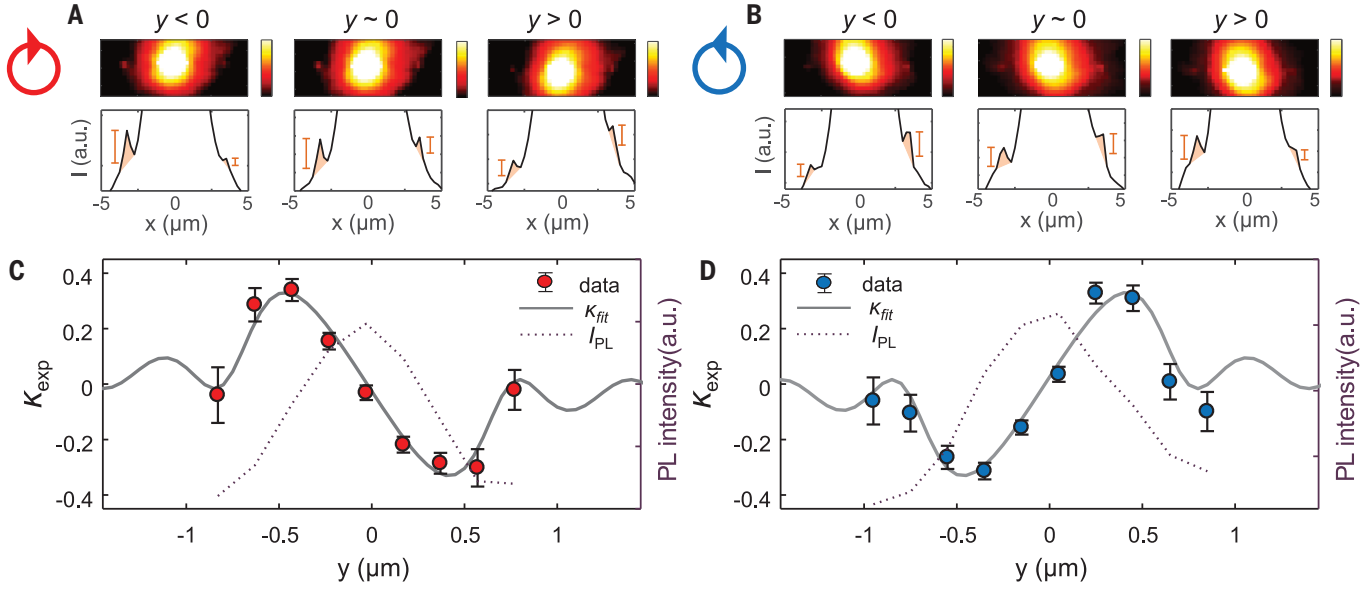


Fig. 3. Experimental demonstration of valley-controlled directional emission. (**A** and **B**) Fluorescence images of the emission of valley-polarized excitons (log scale; top) and line cuts of the intensity profiles (linear scale; bottom) along the silver nanowire under (A) left- and (B) right-handed circularly polarized excitation. The intensities at the excitation regions were deliberately saturated for clarity of the images. The length of the silver nanowire is ~7 μm. (**C** and **D**) Measured directional coupling efficiency, $(I_L - I_R)/(I_L + I_R)$, of the guided emission as a function of the position of the excitation laser with the (C) left- and (D) right-handed circular polarization. Gray lines represent fitting results obtained by using the calculated directional coupling efficiency. Purple dotted lines correspond to the total PL intensity measured from the ends of the nanowire as a function of the excitation position.

with the three-dimensional finite-difference time-domain method. The polarized valley emission is described with a circular dipole. Figure 2D depicts the position-dependent directionality (D_0) of the emission of a left-handed circular dipole in the x-y plane. It is calculated from the light transmitted (T_L and T_R) to the left and right ends of the wire, $D_0(x,y) = [T_L(x,y) - T_R(x,y)]/[T_L(x,y) + T_R(x,y)]$. First, note that the circular dipole emission is preferentially coupled to a propagating mode with an appropriate sign of optical spin (compare Fig. 2, B and C). Second, note that a high degree of directionality is obtained, which is opposite in sign when exciting a circular dipole on either side of the nanowire. A standing wave–like pattern is visible along the wire, which is caused by a small reflection at the end of the finite-length nanowire (7-μm). Note that the high value of directionality of 0.91 is still observed despite the presence of the low-amplitude counterpropagating mode that tends to reduce the t-OSAM and the directionality (*20*). To properly describe the experiment, merely calculating the directionality is insufficient: Actual coupling of the emission to the guided modes is also required; Fig. 2E depicts the normalized, position-dependent coupling strength $\kappa_{tot}(x,y) = [T_L(x,y) + T_R(x,y)]/[T_L(x,y) + T_R(x,y)]_{max}$. Combining D_0 and κ_{tot} results in the chiral-coupling coefficient $\kappa_{valley\text{-}path}$ that describes how well emission from a specific valley couples to a single direction of the light path: $\kappa_{valley\text{-}path}(x,y) = D_0(x,y)\cdot\kappa_{tot}(x,y)$. The calculated $\kappa_{valley\text{-}path}$ (Fig. 2F) shows that the experimentally obtained directionality will be dominated by emitting dipoles located close to the nanowire.

Far-field microscopy is used to verify the chiral valley-photon interface formed by directional coupling of WS_2 excitons to plasmonic nanowires. Valley-polarized excitons are locally excited at the middle of the silver nanowire ($x = 0$) for different y positions with a close-to-diffraction–limited focal spot of a suitably polarized 594-nm laser. Note that the free-space focused laser light cannot couple to the plasmonic guided mode because of momentum mismatch (fig. S3). An optical bandpass filter (620 to 630 nm) is used to predominantly collect the exciton emission for each excitation position. Raw data are presented in Fig. 3, A and B, for left- and right-handed circular excitation polarization, respectively. The large central spot in each image is caused by exciton emission that is not coupled to the plasmonic nanowire. When comparing the images in Fig. 3, A and B, to Fig. 1D, it is clear that the spots at the end of the wire have a lower intensity relative to that of the central excitation spot. This decrease is caused by the fact that Fig. 1D is dominated by the emission from the indirect transition in the near infrared where the propagation losses are considerably less than at 620 to 630 nm, where the direct exciton emission occurs (fig. S3A). In the line traces, the emission coupled to the nanowire shows up as clear peaks (shoulders) on top of a background caused by the excitation spot. In these raw data, the valley-controlled directional emission is already visible to the naked eye: For a given combination of handedness and excitation position, an asymmetry is obvious in the intensity of the spots at either end of the nanowire. The asymmetry of emission is reversed when either the excitation spot is moved to the other side of the nanowire or the polarization handedness of the excitation is flipped.

The PL intensity scattered at the left and right nanowire ends is used to quantify the valley-dependent directionality as a function of y. The extracted $I_L(y)$ and $I_R(y)$, the peak intensities after subtracting the background that arises mainly from the excitation spot, reflect the amount of exciton emission that is coupled to the left- and right-propagating plasmonic modes, respectively (see fig. S4 for the subtracting method). The dotted curves in Fig. 3, C and D, denote the total guided PL intensity $[I_L(y) + I_R(y)]$ as a function of the excitation position. They indicate that the excitons only couple to plasmonic modes when $|y|$ < ~500 nm. We then determine the experimental directional coupling efficiency κ_{exp}: $\kappa_{exp}(y) = [I_L(y) - I_R(y)]/[I_L(y) + I_R(y)]$ (Fig. 3, C and D). We observe a clear directional coupling: For left-handed excitation centered above (y < 0) and below (y > 0) the nanowire, the exciton emission couples to plasmonic eigenstates propagating from right to left and left to right, respectively. When the handedness of the excitation flips, so do the propagation directions of the plasmonic modes. The measurements display exactly the trends observed in the calculations. For symmetric excitation, i.e., excitation on the symmetry axis of the system at the center of the nanowire, κ_{exp} is zero, as expected. A small offset in directionality, caused by a small

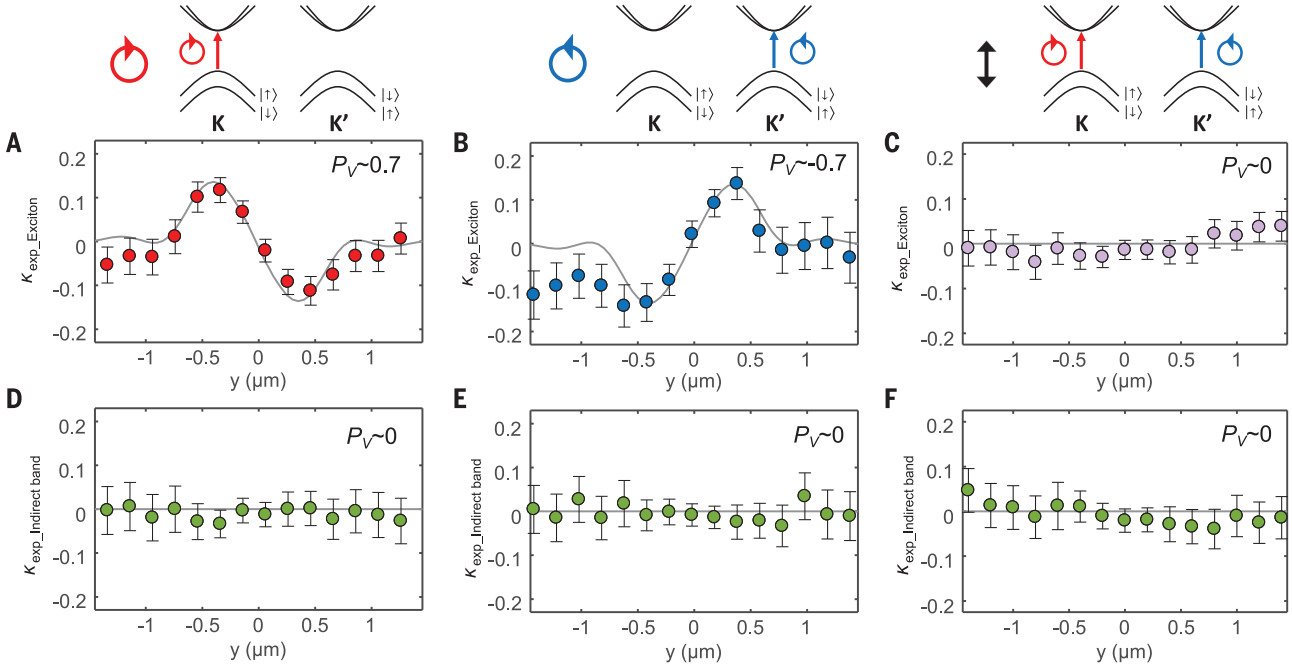


Fig. 4. Direct observation of directional emission induced by valley-path coupling of the valley-polarized exciton. The emission from the valley-polarized excitons shows directional emission, which depends on both the position of excitation and the handedness of the circularly polarized laser (**A** and **B**), whereas the transition of the indirect band gap has a nondirectionality for both handednesses of the circularly polarized laser (**D** and **E**). A linearly polarized laser excitation leads to zero valley polarization so that both the (**C**) exciton and (**F**) indirect-band transitions exhibit no directionality. P_V indicates the degree of valley polarization, $P_V = (I_K - I_{K'})/(I_K + I_{K'})$.

displacement of the excitation spot with respect to the middle of the nanowire in combination with propagation losses, was subtracted from the measured result (fig. S5). A more quantitative description and comparison with calculations will be presented below. The measurements have been reproduced for a number of plasmonic nanowires and WS_2 flakes with different layer thicknesses (fig. S6). In all cases, a strong directionality was observed.

Measurements performed for different excitation polarizations and emission wavelengths confirm that the directional emission is caused by valley-dependent chiral coupling to the plasmonic modes. The direction in the charge-coupled device camera image of the luminescence perpendicular to the length of the nanowire is dispersed in wavelength by using a grating. This configuration provides position-dependent PL spectra along the wire from which we can determine κ_{exp} for different wavelengths (Fig. 4). Figure 4C demonstrates that an equal mix of K and K′ excitons, generated by using a linear polarized excitation light, displays no directionality, whereas for the same nanowire, K or K′ excitons separately (Fig. 4, A and B) display valley-controlled directionality. Note that the maximum κ_{exp} values in Fig. 4, A and B, (~0.17) are lower than the value we observed in Fig. 3 of 0.35. This is caused by the decreased signal-to-noise ratio, resulting from the fact that the wavelength-dispersed image could, by its very nature, not resolve any spatial information along the y direction (fig. S7). Figure 4, D to F, shows that the indirect band gap emission, which does not exhibit valley polarization, does not display any directionality, regardless of position or excitation polarization. Together, these results demonstrate that all experimentally observed directionality is attributable to valley-controlled chiral coupling to plasmonic modes.

Comparison of the experimental and the calculated results reveals two differences. First, the experimental directionality profile is spatially smeared out. Second, the magnitude of κ_{exp} is less than maximum in the calculations. This is caused, in part, by obvious experimental limitations, e.g., a finite excitation spot size and background noise: Both decrease the experimental maximum of the directionality, and the finite-excitation spot size also spatially smears out the pattern. However, the reduction in directionality is also caused by the finite P_V of the WS_2 system itself. The valley-to-valley hopping (*29*) would actually flip the direction in which the plasmonic modes are launched. Therefore, the effective directionality with a finite P_V would be a simple product of $\kappa_{valley\text{-}path}$ and P_V. These experimental matters are readily taken into account to properly compare our realized chiral valley-photon interface to the ideal simulated interface

$$\kappa_{fit}(y) = \frac{[\kappa_{valley-path} \cdot P_V] \otimes PSF}{[\kappa_{tot} + 2\rho] \otimes PSF}$$

where $\kappa_{valley\text{-}path}$ is the calculated chiral valley-path coupling coefficient for a perfect circularly polarized dipole, P_V is the measured degree of valley polarization (0.7), and ρ is an unpolarized background noise relative to the maximum signal that represents a heuristic decrease of chiral coupling. To account for the effect of the finite size of the excitation region, we convolute the calculated signal with the point spread function estimated from the experimental data (figs. S8 and S9). We obtain excellent agreement between fit and data for $\rho = 0.009 \pm 0.005$ (gray lines in Fig. 3, C and D). With a reduction of directionality as $D_0/(1 + 2\rho/\kappa_{tot})$ the fitting procedure reveals that the experimental valley-to-path coupling efficiency, including background noise, is as high as 0.90 ± 0.01. The quality of the fit yields a number of important conclusions. First, because the experimentally determined point spread function is sufficient to explain the broadening of the directional coupling efficiency, exciton diffusion does not notably affect the experiment. Second, any effect of the plasmonic nanowire on the polarization of the excitation focal spot is negligible. Lastly, and most importantly, the chiral valley-to-path coupling is only limited by the magnitude of the transverse optical spin of the plasmonic nanowire modes. Thus, we have been able to realize a room-temperature interface between the valley pseudospin of WS_2 and the propagation direction of nanowire plasmonic modes with a fidelity as high as 0.9.

Our results show that the efficient coupling of valley-polarized excitons in WS_2 to the transverse optical spin of plasmonic nanowire modes

results in valley-photon direction locking with high fidelity. The high valley-spin coupling strength of WS_2 and high transverse optical spin density in plasmonic modes provide a chiral valley (spin)–photon interface at room temperature and without the necessity of external magnetic fields. It is important to note that because of the one-to-one relation between optical path and the local transverse optical spin, the propagation direction of the guided light can also be exploited to transfer the valley degree of freedom to other valley devices on a chip through suitable, localized breaking of mirror symmetry, again exploiting the transverse optical spin of the mode, but this time to excite. Realization of such an interface presents a platform for both fundamental studies and a wide range of exciting applications of chiral photonics and chiral quantum optics.

REFERENCES AND NOTES

1. D. Xiao, W. Yao, Q. Niu, *Phys. Rev. Lett.* **99**, 236809 (2007).
2. J. R. Schaibley *et al.*, *Nat. Rev. Mater.* **1**, 16055 (2016).
3. K. F. Mak, J. Shan, *Nat. Photonics* **10**, 216–226 (2016).
4. Q. H. Wang, K. Kalantar-Zadeh, A. Kis, J. N. Coleman, M. S. Strano, *Nat. Nanotechnol.* **7**, 699–712 (2012).
5. K. F. Mak, K. L. McGill, J. Park, P. L. McEuen, *Science* **344**, 1489–1492 (2014).
6. Y. J. Zhang, T. Oka, R. Suzuki, J. T. Ye, Y. Iwasa, *Science* **344**, 725–728 (2014).
7. A. M. Jones *et al.*, *Nat. Nanotechnol.* **8**, 634–638 (2013).
8. X. Xu, W. Yao, D. Xiao, T. F. Heinz, *Nat. Phys.* **10**, 343–350 (2014).
9. D. Xiao, G.-B. Liu, W. Feng, X. Xu, W. Yao, *Phys. Rev. Lett.* **108**, 196802 (2012).
10. H. Zeng, J. Dai, W. Yao, D. Xiao, X. Cui, *Nat. Nanotechnol.* **7**, 490–493 (2012).
11. K. F. Mak, K. He, J. Shan, T. F. Heinz, *Nat. Nanotechnol.* **7**, 494–498 (2012).
12. T. Cao *et al.*, *Nat. Commun.* **3**, 887 (2012).
13. J. Petersen, J. Volz, A. Rauschenbeutel, *Science* **346**, 67–71 (2014).
14. C. Sayrin *et al.*, *Phys. Rev. X* **5**, 041036 (2015).
15. R. Mitsch, C. Sayrin, B. Albrecht, P. Schneeweiss, A. Rauschenbeutel, *Nat. Commun.* **5**, 5713 (2014).
16. F. J. Rodríguez-Fortuño *et al.*, *Science* **340**, 328–330 (2013).
17. D. O'Connor, P. Ginzburg, F. J. Rodríguez-Fortuño, G. A. Wurtz, A. V. Zayats, *Nat. Commun.* **5**, 5327 (2014).
18. K. Y. Bliokh, D. Smirnova, F. Nori, *Science* **348**, 1448–1451 (2015).
19. K. Bliokh, F. Rodríguez-Fortuño, F. Nori, A. V. Zayats, *Nat. Photonics* **9**, 796–808 (2015).
20. R. J. Coles *et al.*, *Nat. Commun.* **7**, 11183 (2016).
21. B. le Feber, N. Rotenberg, L. Kuipers, *Nat. Commun.* **6**, 6695 (2015).
22. F. J. Rodríguez-Fortuño, I. Barber-Sanz, D. Puerto, A. Griol, A. Martínez, *ACS Photonics* **1**, 762–767 (2014).
23. K. Y. Bliokh, A. Y. Bekshaev, F. Nori, *Nat. Commun.* **5**, 3300 (2014).
24. A. Aiello, P. Banzer, M. Neugebauer, G. Leuchs, *Nat. Photonics* **9**, 789–795 (2015).
25. T. Van Mechelen, Z. Jacob, *Optica* **3**, 118–126 (2016).
26. T. Chervy, S. Azzini, E. Lorchat, S. Wang, Y. Gorodetski, J. A. Hutchison, S. Berciaud, T. W. Ebbesen, C. Genet, arXiv:1701.07972 [cond-mat.mtrl-sci] (27 January 2017).
27. B. Zhu, H. Zeng, J. Dai, Z. Gong, X. Cui, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 11606–11611 (2014).
28. A. M. Jones *et al.*, *Nat. Phys.* **10**, 130–134 (2014).
29. R. Schmidt *et al.*, *Nano Lett.* **16**, 2945–2950 (2016).

ACKNOWLEDGMENTS

S.-H.G., F.A., and L.K. acknowledge funding from European Research Council (ERC) Advanced Investigator Grant no. 340438-CONSTANS. This work is part of the research program of The Netherlands Organization for Scientific Research (NWO). F.A. acknowledges support from the Marie Skłodowska-Curie individual fellowship BISTRO-LIGHT (no. 748950). B.S. and E.C.G. were supported by funding from the ERC under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC grant agreement no. 337328, "NanoEnabledPV." All data are reported in the main text and supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6374/443/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S14
References (*30*–*35*)

15 September 2017; accepted 12 December 2017
10.1126/science.aan8010

LIGHT METALS

Mechanistic origin and prediction of enhanced ductility in magnesium alloys

Zhaoxuan Wu,[1,2] Rasool Ahmad,[1] Binglun Yin,[1] Stefanie Sandlöbes,[3] W. A. Curtin[1]*

Pure magnesium exhibits poor ductility owing to pyramidal $\langle c + a \rangle$ dislocation transformations to immobile structures, making this lowest-density structural metal unusable for many applications where it could enhance energy efficiency. We show why magnesium can be made ductile by specific dilute solute additions, which increase the $\langle c + a \rangle$ cross-slip and multiplication rates to levels much faster than the deleterious $\langle c + a \rangle$ transformation, enabling both favorable texture during processing and continued plastic straining during deformation. A quantitative theory establishes the conditions for ductility as a function of alloy composition in very good agreement with experiments on many existing magnesium alloys, and the solute-enhanced cross-slip mechanism is confirmed by transmission electron microscopy observations in magnesium-yttrium. The mechanistic theory can quickly screen for alloy compositions favoring conditions for high ductility and may help in the development of high-formability magnesium alloys.

Developing high-performance, affordable, lightweight structural metals is an important goal for achieving energy efficiency, safety, and human well-being. Mg is the lightest structural metal (about two-thirds and one-fourth the densities of Al and Fe, respectively), abundant in Earth, recyclable, and biocompatible. These properties make Mg attractive for automotive, aerospace, and biomedical applications. However, Mg has low ductility, making it difficult to process at room temperature and preventing its use in many applications. The poor properties of Mg are connected to its hexagonal close-packed (hcp) crystal structure. Plastic slip in the crystallographic *c* direction is necessary for generalized plasticity, but the required easy-glide pyramidal $\langle \mathbf{c} + \mathbf{a} \rangle$ dislocations undergo a rapid transition to an immobile structure (*1*) that limits *c*-axis plastic strain. This is exacerbated by a difference of a factor of ~100 in critical resolved shear stress (CRSS) for the $\langle \mathbf{c} + \mathbf{a} \rangle$ slip as compared with the basal $\langle \mathbf{a} \rangle$ slip. Metallurgical strategies (alloying and thermomechanical processing) aim to increase ductility by engineering grain size, randomizing texture away from the unfavorable strong basal texture in sheet forming, strengthening basal slip, or activating prism slip or twinning (Fig. 1). Rare earth (RE) solutes (Y, Tb, Dy, Ho, Er, Ce, and Gd) (*2*–*4*) at very low concentrations [~0.03 to 1.0 atomic % (at %)] stand out as yielding good room-temperature ductility even at fairly large grain sizes and moderately strong basal texture. However, the mechanistic origins of the enhanced ductility are unknown (*5*), so the creation of new ductile non-RE Mg alloys is largely empirical.

We present a physical mechanism and associated theory for achieving enhanced ductility in Mg alloys. Specifically, appropriate solid-solution alloying can greatly increase the rate of $\langle \mathbf{c} + \mathbf{a} \rangle$ screw dislocation cross-slip. Cross-slip and double cross-slip then naturally lead to multiplication of the easy-glide $\langle \mathbf{c} + \mathbf{a} \rangle$ dislocations at rates much faster than the deleterious transformations of the $\langle \mathbf{c} + \mathbf{a} \rangle$ dislocations (Fig. 2). This reduces the effective $\langle \mathbf{c} + \mathbf{a} \rangle$ CRSS toward that of the easy-glide pyramidal configuration (*1*) and provides the *c*-axis strain accommodation that satisfies the Von Mises criteria, thus enabling high ductility during deformation. Solute-enhanced $\langle \mathbf{c} + \mathbf{a} \rangle$ cross-slip and slip also contribute to thermomechanical processing of favorable textures by helping to rotate grains into crystallographic orientations that weaken the basal texture normally caused by dominant basal $\langle \mathbf{a} \rangle$ slip and extension twinning (*6*). In particular, $\langle \mathbf{c} + \mathbf{a} \rangle$ slip tends to result in a DPRD (double-peak basal pole tilted in the rolling direction) texture during rolling and plane-strain compression (*6*). Enhanced $\langle \mathbf{c} + \mathbf{a} \rangle$ slip can also create three-dimensional dislocation networks that drive recrystallization, refine grains, and help weaken texture. Not withstanding other mechanisms (*7*, *8*), solute-accelerated $\langle \mathbf{c} + \mathbf{a} \rangle$ cross-slip and activity can create favorable conditions for high ductility, with ductility then limited by normal work-hardening mechanisms.

A mechanistic, quantitative theory establishes the conditions enabling high ductility as a function of alloy composition. Predictions using density functional theory (DFT)–computed inputs then resolve the puzzle of how RE, Zr, Ca, and Mn solutes can induce ductility at extremely low concentrations (c ~ 0.03 to 0.3 at %) and rationalize the ductility trends across a wide range of existing alloys. Transmission electron microscopy (TEM) studies in Mg and Mg-Y further show explicitly that the alloys have enhanced $\langle \mathbf{c} + \mathbf{a} \rangle$ cross-slip, enhanced $\langle \mathbf{c} + \mathbf{a} \rangle$ activity, and substantially higher ductility relative to pure Mg with a similar texture.

In pure Mg, the pyramidal-to-basal (PB) transition of the edge pyramidal II $\langle \mathbf{c} + \mathbf{a} \rangle$ dislocation has an activation energy barrier ΔE_{PB} of only ~0.5 eV (*1*). Cross-slip (and double cross-slip) of $\langle \mathbf{c} + \mathbf{a} \rangle$ screw dislocations, leading to new dislocation loops that expand and generate plasticity, can only effectively circumvent this transition if the cross-slip activation energy ΔG_{XS} is much lower than ΔE_{PB}. Pyramidal dislocation cross-slip in hcp Mg requires that the screw dislocations move from the lower-energy pyramidal II slip planes to the higher-energy pyramidal I planes (Fig. 2). While the intrinsic cross-slip activation energy is only ~0.25 eV, an additional energy scaling with cross-slip nucleation length is required owing to the small (but crucial) difference in energy per unit length $\Delta E^{\mathrm{I-II}}$ between pyramidal I and II screw dislocations (*9*). In pure Mg, the additional energy $\Delta E^{\mathrm{I-II}}_{\mathrm{Mg}}$ greatly increases the total cross-slip activation energy ΔG_{XS} so that it exceeds ΔE_{PB}, making cross-slip ineffective at circumventing the detrimental effects of the PB transition. With appropriate alloying, $\Delta E^{\mathrm{I-II}}$ can be reduced so that ΔG_{XS} is much lower than ΔE_{PB}, enabling rapid cross-slip, dislocation multiplication, and ultimately greater ductility at similar grain sizes and textures.

To formalize the mechanism, we consider a dislocation loop of $L \times L$ (Fig. 2). Enhanced ductility is achieved when the total $\langle \mathbf{c} + \mathbf{a} \rangle$ cross-slip rate R_{XS}, owing to possible nucleation at L/l_{XS} segments along the screw section of length L, is much faster than the total $\langle \mathbf{c} + \mathbf{a} \rangle$ PB transition rate R_{PB}, owing to possible nucleation at L/l_{PB} segments along the edge section of length L. Here, l_{XS} and l_{PB} denote the critical nucleation lengths for these two thermally activated processes. Effective cross-slip for high ductility then requires that

$$\nu_0 (L/l_{\mathrm{XS}}) e^{-\Delta G_{\mathrm{XS}}/(kT)} \gg \nu_0 (L/l_{\mathrm{PB}}) e^{-\Delta E_{\mathrm{PB}}/(kT)} \quad (1)$$

where $l_{\mathrm{PB}} \approx 2$ nm (*1*) and ν_0, k, and T are the attempt frequency, Boltzmann constant, and temperature, respectively. To quantify conditions for achieving high ductility, we introduce the "ductility index" χ, where 10^{χ} is the factor by which the cross-slip rate exceeds the PB transition rate. From Eq. 1, for ductility at level χ, the cross-slip barrier should satisfy

$$\Delta G_{\mathrm{XS}} = \Delta E_{\mathrm{PB}} - kT \ln(10^{\chi}\, l_{\mathrm{XS}}/l_{\mathrm{PB}}) \quad (2)$$

We consider favorable conditions for ductility as corresponding to $\chi > 1$ (cross-slip 10 times

[1]Institute of Mechanical Engineering, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne CH-1015, Switzerland. [2]Institute of High Performance Computing, 1 Fusionopolis Way, #16-16 Connexis, 138632, Singapore. [3]Institut für Metallkunde und Metallphysik, RWTH (Rheinisch-Westfälische Technische Hochschule) Aachen University, Kopernikusstraße 14, 52074 Aachen, Germany.
*Corresponding author. Email: william.curtin@epfl.ch

as fast as the PB transition) and poor conditions for ductility as corresponding to $\chi < 0$ (cross-slip slower than the PB transition). The basal-transformed pyramidal II dislocation is so energetically favorable (~0.3 eV/Å) that changing ΔE_{PB} is unlikely (*1*, *5*). So, increasing ductility is achieved by reducing the cross-slip barrier ΔG_{XS}.

The cross-slip process from the low-energy pyramidal II plane to the high-energy pyramidal I plane is driven by the net resolved shear stress $\Delta\tau$ acting to bow out any nucleating cross-slip segment on the pyramidal I plane. We write the cross-slip energy barrier ΔG_{XS} as (*9*)

$$\Delta G_{XS} = \Delta G_{XS,i} + \Delta E^{I-II} l_{XS} + \Gamma \Delta s - \Delta\tau b A \quad (3)$$

The intrinsic cross-slip barrier $\Delta G_{XS,i}$ is associated with the nucleation of dislocation cross-slip jogs of width $l_{nuc} \approx 2.5$ nm for $\langle \mathbf{c}+\mathbf{a} \rangle$ cross-slip in pure Mg (*9*); $\Delta E^{I-II} l_{XS}$ is the extra energy cost to create the pyramidal I segment of length $l_{XS} = l_{nuc} + l_{CXS}$ (fig. S3); $l_{CXS} = 2\sqrt{\Delta E^{I-II}(2\Gamma - \Delta E^{I-II})}/(b\Delta\tau)$ is the critical bow-out length, with Γ being the pyramidal I line tension; and the last two terms are the energy to create the additional bowed-out dislocation length Δs on the pyramidal I plane, minus the work done by $\Delta\tau$ over the bowed-out area A. Δs and A are fully determined by l_{CXS}, Γ, b (the magnitude of the Burgers vector), and $\Delta\tau$ (*10*). The cross-slip barrier ΔG_{XS} can be reduced by reducing ΔE^{I-II} through alloying.

In alloys, ΔE^{I-II} has a contribution that is linear in solute concentration c owing to the difference in solute interactions with the pyramidal I and II stacking faults (SFs) ΔE^{I-II}_{SF}, which is computed using DFT (*10*). There is also a contribution due to the solute misfit strain interaction with the dislocation stress field, but this term cannot be separated from the SF interaction because there are high stresses in the SF. Fortunately, the total misfit interaction is small compared with ΔE^{I-II}_{SF} in all solutes, except Al. In Al, ΔE^{I-II}_{SF} is small, and the solute misfit energies in the SF are nearly identical to the first-principles values (*10*), so we use the total misfit energy difference for Al only.

We show the predicted $\Delta E^{I-II}(c)$ relative to pure Mg versus c for a wide range of solutes (Fig. 3A). The RE solutes (Ce, Gd, Nd, Y, and Er), as well as Zr and Ca, stand out as especially effective in decreasing the average pyramidal I–II energy difference and so can enhance cross-slip and ductility. Mn has a smaller effect but can still enhance cross-slip. In contrast, Zn and Ag increase $\Delta E^{I-II}(c)$ considerably and so are counterproductive to ductility, on average. Overall, the trends in average solute effects on the pyramidal I–II energy difference are qualitatively consistent with the trends in ductility versus solute from experiments (Fig. 1), correctly identifying favorable and unfavorable solutes.

Solute fluctuations exist naturally in random alloys. Favorable statistical distributions of solutes over the critical length l_{XS} given above reduce the cross-slip barrier. The standard deviation

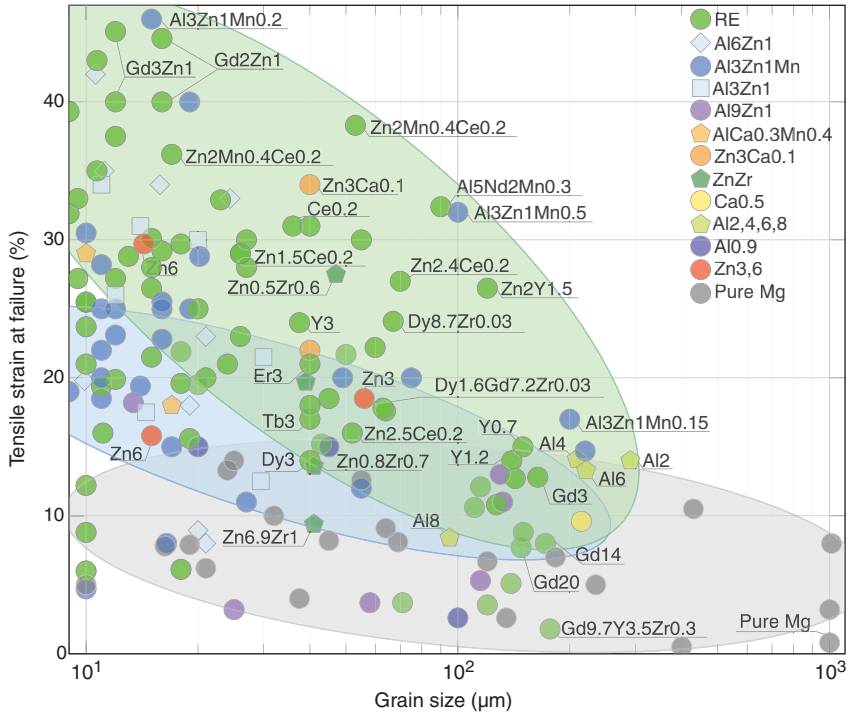


Fig. 1. Room-temperature tensile failure strain, as a measure of ductility, versus grain size in selected polycrystalline Mg and Mg alloys. In general, ductility is very low in pure Mg (gray ellipse) and increases moderately (blue ellipse; mainly Mg-Al-Zn and Mg-Al-Zn-Mn) and substantially (green ellipse; mainly Mg-RE) with alloying, while generally increasing with decreasing grain size. The RE alloys (Y, Er, Tb, and Dy) at 3 wt % (~1 at % for Y and 0.3 at % for others), 0.2 wt % Ce, and 0.1 wt % Ca have comparatively high ductility for a given grain size. Labels indicate the weight percentage of each solute. Variations in ductility, especially points outside the ellipses, are attributed to specific or selected loading directions with respect to special crystallographic textures. A much larger and fully referenced data set showing the same trends is provided in fig. S1. All data are from experiments at nominally quasi-static strain rates.

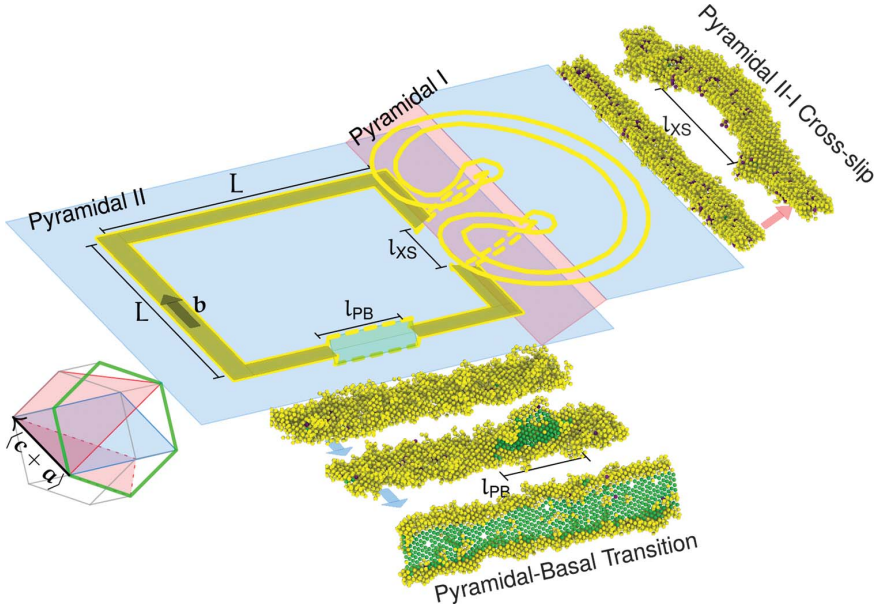


Fig. 2. Competing pyramidal-to-basal transition and pyramidal II–I cross-slip processes during the expansion of an *L* × *L* ⟨c + a⟩ dislocation loop on the pyramidal II plane. l_{PB} and l_{XS} denote the critical nucleation lengths of the two thermally activated processes. Spheres are atoms colored according to their local atomic environment as identified by the common neighbor analysis (green, face-centered cubic; purple, body-centered cubic; yellow, others); only non-hcp atoms are shown.

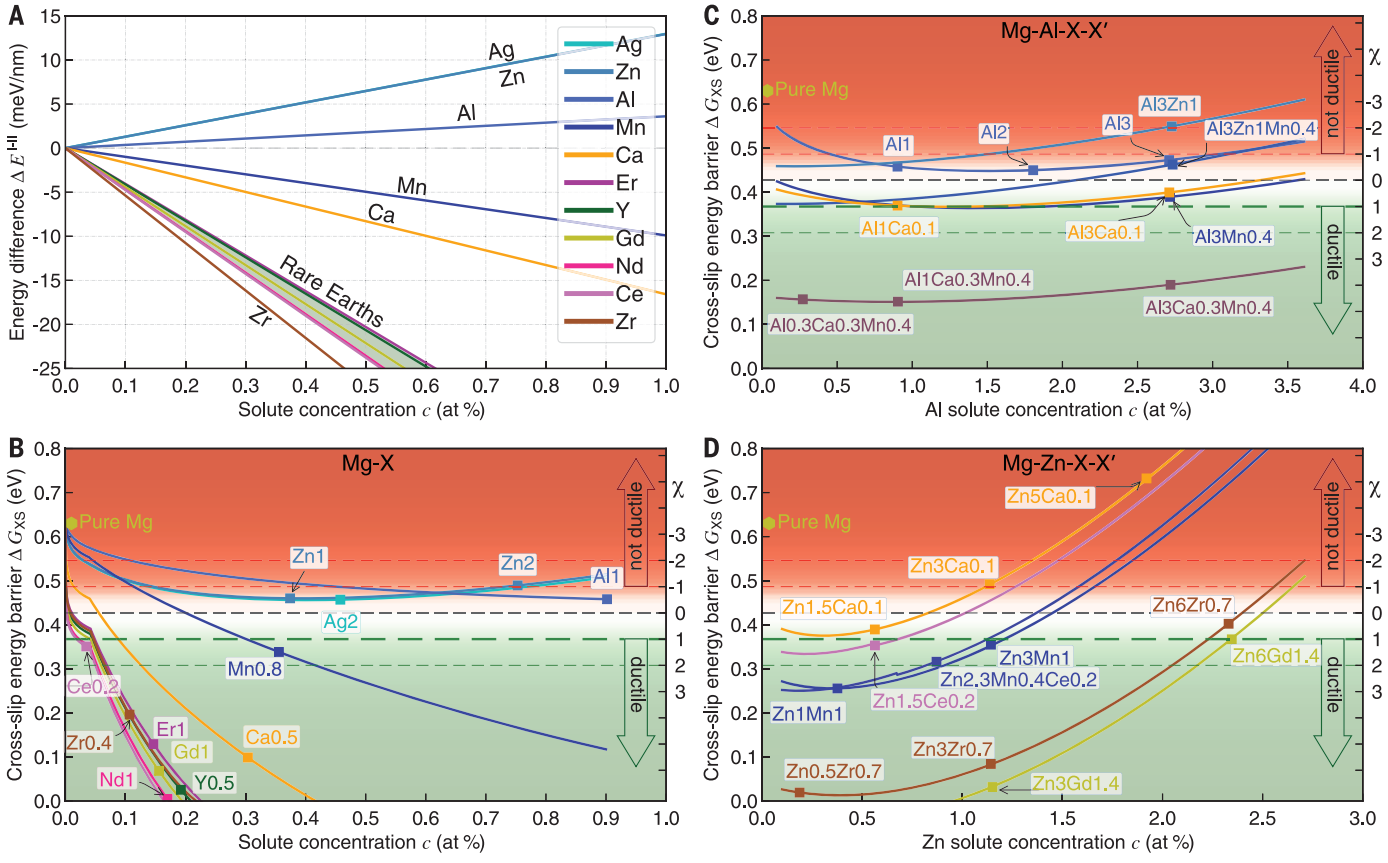


Fig. 3. Activation energy for cross-slip and ductility index χ for binary and higher-order Mg alloys. (**A**) Average solute contribution to energy difference $\Delta E^{I-II}(c)$ between the pyramidal I and II $\langle \mathbf{c}+\mathbf{a} \rangle$ screw dislocations for various REs, Al, Zn, Zr, Ca, Mn, and Ag, immediately showing which solutes will be effective in enhancing ductility. (**B**) Predicted pyramidal II–I cross-slip activation energy barrier including solute fluctuations and ductility index χ for binary Mg alloys as a function of solute concentration c for the same solutes. $\chi > 1$ indicates favorable conditions for ductility. RE solutes achieve $\chi > 1$ at very low concentrations; Zr and Ca are also highly effective, and Mn is moderately effective. Zn and Ag (almost identical) have $\chi < 0$ and do not reach favorable conditions for ductility. (**C**) Mg-Al-X-X′ with varying Al concentrations and (**D**) Mg-Zn-X-X′ with varying Zn concentrations. In (C), Al-Ca, Al-Mn, and Al-Ca-Mn reach the favorable ductility condition $\chi > 1$ over some ranges of Al concentrations, and binary Mg-Al approaches $\chi = 0$ at 1 to 3 at % Al. In (D), Zn-Ce, Zn-Mn, Zn-Mn-Ce, Zn-Zr, and Zn-Gd reach the favorable ductility condition $\chi > 1$ at low Zn concentrations, and χ decreases as the Zn concentration increases. In (C) and (D), predictions are shown for Al and Zn concentrations above their very dilute limits; 0.1 wt % Ca and 0.2 wt % Ce can be in the very dilute limits. All predictions are consistent with the trends shown in Fig. 1. Attainable solute concentrations are limited by solubility and precipitation, factors not assessed here. The individual labels indicate solute weight percentage to make contact with standard alloy nomenclature.

of the statistical distribution of the solute-SF interaction energy difference scales as ΔE^{I-II}_{fluc}, which is computed over all unique solute sites within a unit length b (*10*). The typical favorable (energy-lowering) solute fluctuations within l_{XS} then lead to a solute-modified cross-slip energy barrier

$$\Delta G_{XS} = (\Delta G_{XS,i} + \Delta E^{I-II}_{Mg} l_{XS} + \Gamma \Delta s - \Delta\tau b A) + (\Delta E^{I-II}_{SF} l_{XS} c - \Delta E^{I-II}_{fluc} \sqrt{l_{XS} b c}) \quad (4)$$

In Eq. 4, the first set of terms is ΔG_{XS} for pure Mg (Eq. 3), the second term is due to the average effects of solutes on the SF energies, and the third term accounts for the spatial fluctuations in the random solute distribution. At very low concentrations $c < c^* \approx b/(50 l_{XS})$, a discrete statistical analysis is required (*10*). We compute the ductility index χ versus alloy composition by combining Eqs. 2 and 4. Compositions yielding $\chi > 1$ and $\chi < 0$ respectively correspond to favorable and unfavorable conditions for achieving high ductility but do not directly predict measured ductility. Below, we correlate χ with experimental texture and tensile ductility.

We show the effective cross-slip barrier ΔG_{XS} and ductility index χ versus solute concentration at T = 300 K for many binary alloys (Figs. 1 and 3B). We use $\Delta E^{I-II}_{Mg} = 17$ meV/nm, which is the average of DFT and modified embedded atom method estimates, and $\Delta\tau$ = 10 MPa, which is related to the difference between pyramidal II and I Peierls stresses (*10*). The model predicts that all RE (Ce, Y, Nd, Gd, and Er) alloys reach $\chi > 1$ at dilute concentrations. For Ce, $c \approx 0.035$ at % is sufficient, consistent with the observed very weak texture (*11*, *12*) and high ductility (*12*, *13*). For other REs, results are consistent with measured weak to moderate DPRD texture and high ductility (*3*, *4*, *11*), and they rationalize why grain refinement, texture weakening, and ductility are more sensitive to atomic concentration than to RE type (*4*, *11*, *14*). Zr and Ca are also predicted to be very effective at levels below the solubility limits [e.g., $c \approx 0.17$ at % for Zr (*15*, *16*)]. Ca results are consistent with moderate ductility at 0.036 and 0.3 at % Ca in large-grain-size alloys (200 to 400 μm) (*17*, *18*) and with the observed DPRD texture (*19*). For Mn, the ductility condition is reached at $c \approx 0.3$ at %, in agreement with observed ductility and formability at 0.35 at % and 18-μm grain size (*20*). For Zn and Ag, $\chi < 0$ at all concentrations, so these solutes are not predicted to create favorable conditions for ductility or for DPRD texture, consistent with experiments (*19*). Al has weak effects with a minimum $\chi \approx -0.3$ at 2.0 at % (Fig. 3C),

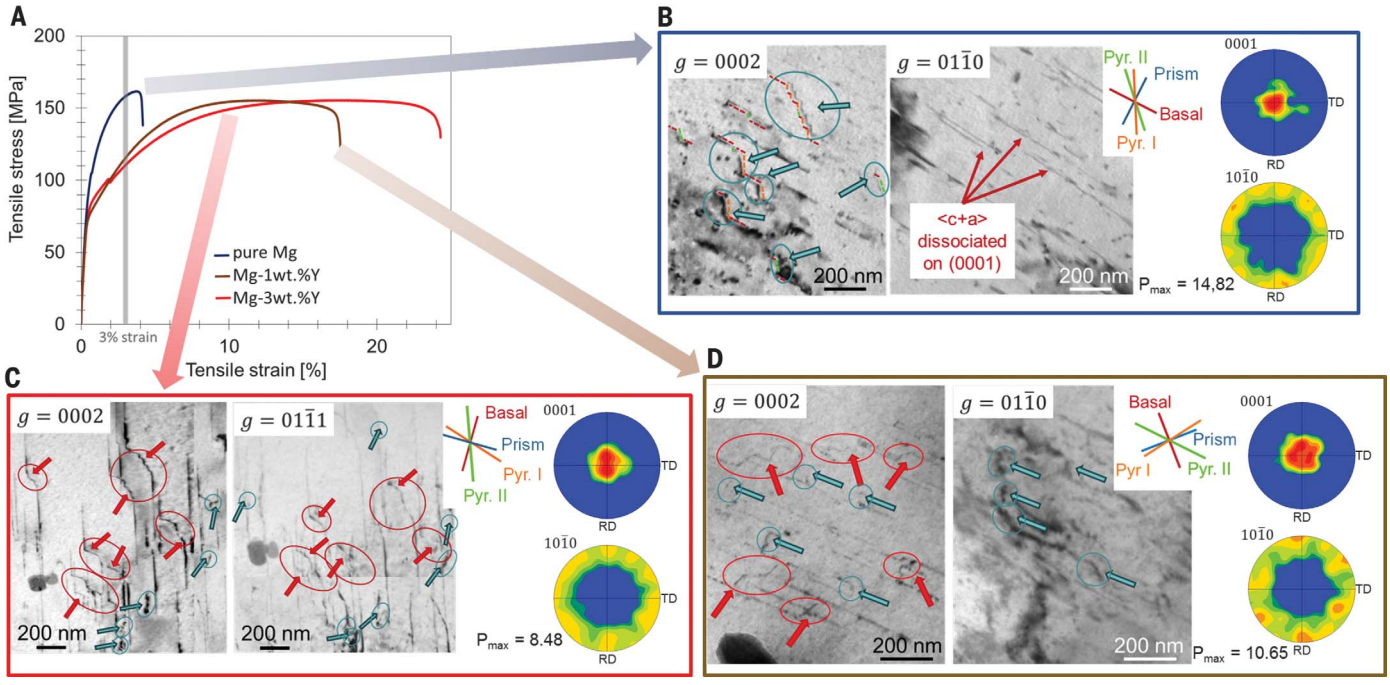


Fig. 4. Experimental results for pure Mg and Mg-Y alloys under tensile deformation at room temperature. (**A**) Stress-strain curves and (**B** to **D**) initial {0001} and {10$\bar{1}$0} pole figures and bright-field TEM of the corresponding dislocation microstructures after 2.5 to 3% strain for pure Mg (B), Mg–3 wt % Y (C), and Mg–1 wt% Y (D). In (B) to (D), dislocations visible under diffraction vector **g** = 0002 are $\langle \mathbf{c}+\mathbf{a} \rangle$ dislocations based on two-beam $\mathbf{g}\cdot\mathbf{b}$ = 0 visibility analysis (**b**, Burgers vector); the insets show the corresponding crystallographic plane traces identified using diffraction analysis, enabling determination of the crystal planes of the observed $\langle \mathbf{c}+\mathbf{a} \rangle$ dislocations. In (B), $\langle \mathbf{c}+\mathbf{a} \rangle$ dislocation segments are lying on basal and pyramidal II planes; areas where dislocations change from pyramidal II to basal planes are highlighted by turquoise arrows and circles. In (C) and (D), $\langle \mathbf{c}+\mathbf{a} \rangle$ dislocation segments are predominantly on pyramidal II planes, with frequent plane changes to either pyramidal I (highlighted by red arrows and circles) or basal (highlighted by turquoise arrows and circles) planes. In (C), for Mg–3 wt % Y, high-frequency $\langle \mathbf{c}+\mathbf{a} \rangle$ dislocation switching between pyramidal II and I planes (double cross-slip) is evident.

barely lower than the minima for Zn or Ag, whereas experiments show moderate ductility in solution-treated Mg-Al (*21*) and DPRD texture in as-rolled Mg-Al (*19*). Precise results are sensitive to ΔE_{Mg}^{I-II} and $\Delta\tau$ (*10*), so conclusions about ductility for cases where $\chi \sim 0$ are less definitive. Nonetheless, the overall predicted trends of the ductility and texture in binary alloys are in broad agreement with the trends in Fig. 1 and fig. S1.

We now show direct physical evidence for the operation of the pyramidal $\langle \mathbf{c}+\mathbf{a} \rangle$ cross-slip mechanism. We show the initial {0001} and {10$\bar{1}$0} pole figures, stress-strain curves, and dislocation microstructures in pure Mg, Mg–1 weight % (wt %) Y (0.28 at % Y), and Mg–3 wt % Y (0.84 at % Y) alloys deformed under uniaxial tensile loading at room temperature (Fig. 4). All samples are process-controlled to yield similar grain sizes and the strong basal texture typical of pure Mg (*10*), with all TEM observations performed in grains having close-to-basal orientation as confirmed by selected area diffraction, making direct comparisons appropriate. Upon yielding, the alloys have much lower hardening rates that could be due to both the slightly weaker texture (favoring more basal slip) and enhanced $\langle \mathbf{c}+\mathbf{a} \rangle$ slip. We thus examine the deformation microstructures at the same ~3% strain (Fig. 4, B to D). The stress level for pure Mg is much higher than for the two alloys. The alloys exhibit a clear increase in both $\langle \mathbf{c}+\mathbf{a} \rangle$ dislocation cross-slip and overall $\langle \mathbf{c}+\mathbf{a} \rangle$ activity: Connected $\langle \mathbf{c}+\mathbf{a} \rangle$ dislocation segments are increasingly found on both pyramidal I and II planes with increasing Y concentrations, with pyramidal II slip being dominant. This is consistent with earlier observations of slip traces on both pyramidal I and II planes in Mg–3 wt % Y (*22*). In pure Mg, the $\langle \mathbf{c}+\mathbf{a} \rangle$ dislocations are primarily dissociated on the basal plane, with some indications of limited cross-slip involving either pyramidal II or I slip. These observations indicate that, in the alloys, $\langle \mathbf{c}+\mathbf{a} \rangle$ slip operates and contributes to the lower hardening even at the very early stages of plastic straining, consistent with the theory that posits continued $\langle \mathbf{c}+\mathbf{a} \rangle$ generation and slip at lower stresses. We attribute the presence of some pyramidal I segments in pure Mg to the high tensile stress level in pure Mg (*10*), which is consistent with single-crystal studies (*23*) and atomistic simulations (*9*). The role of easy $\langle \mathbf{c}+\mathbf{a} \rangle$ slip is perhaps more critical at the later stages of deformation (strain > 10% where further extension twinning is limited). In pure Mg polycrystals with grain sizes > 10 μm, excessive basal slip without *c*-axis strain accommodation can quickly build up strain incompatibility and high constraint stresses, leading to damage initiation and limiting ductility (*24*). In the Mg-Y alloys, the stress-strain curves clearly have a longer range of stable plastic flow at relatively normal hardening rates, indicating that sufficient plastic deformation mechanisms are activated until the onset of macroscopic plastic instability (e.g., the Considère criterion) is reached. These controlled experiments demonstrate that Y solutes do enhance $\langle \mathbf{c}+\mathbf{a} \rangle$ dislocation cross-slip, slip, *c*-axis strain accommodation, and ductility with increasing Y concentration, in excellent agreement with our predictions.

The theory is applicable to dilute solid-solution ternary alloys and other materials. We show the predicted trends for ternary and quaternary alloys involving Al and Zn, respectively, corresponding to many experimental alloys (Figs. 1 and 3, C and D). For Mg-Al-Zn (commercial AZ alloys), the theory predicts $\chi < -1$ at the widely used 1 wt % Zn (e.g., AZ31 alloy with 0.3 at % Zn), indicating poor ductility conditions and thus likely requiring special processing to achieve nonbasal textures. This may account for the very wide range of reported failure strains and formability [Fig. 1 and (*25*)]. The addition of just 0.4 wt % Mn to create the AZ31B alloy yields $\chi \sim 0$, largely counteracting the deleterious effects of Zn, and this is consistent with the trend of increased ductility for this slightly modified alloy as compared

with commercial AZ31 alloy. Mg-Al-Ca at only 0.06 at % Ca with 1 at % Al attains the ductility condition (χ = 1), consistent with the DPRD texture and good ductility in experiments (*19*), whereas 1 at % Al alone has χ < 0. The Mg–Al–0.4 wt % Mn alloys are predicted to be nearly identical to the Mg–Al–0.1 wt % Ca alloys. Experiments show that extruded and hot-rolled Mg–(3 to 8) Al–0.4 Mn (weight %) alloys have weak texture with DPRD features and high tensile failure strain (*25*), consistent with enhanced $\langle \mathbf{c}+\mathbf{a} \rangle$ slip owing to Al and Mn solutes. Last, the recently reported Mg-Al-Ca-Mn system at 0.18 at % Ca and 0.17 at % Mn (*26*, *27*) is predicted to have χ > 2 across a wide range of Al concentrations, and indeed these alloys demonstrate good tensile ductility.

For Mg-Zn-X-X′ alloys, high Zn concentrations raise the cross-slip barrier considerably, leading to χ < 0, which is unfavorable for ductility. This is consistent with broad experimental trends where ductility generally decreases or saturates with increasing Zn (*19*, *28*–*32*) (fig. S1). The deleterious effect of Zn also corroborates the texture evolution during rolling, where the DPRD texture, driven by $\langle \mathbf{c}+\mathbf{a} \rangle$ slip in Mg-Ce (*33*, *34*), Mg-Y (*35*), and Mg-Ca (*19*, *36*), weakens or disappears upon additions of Zn (*13*, *19*, *37*). However, alloying with dilute RE solutes can easily yield χ > 1 at low Zn concentrations. We predict that Mg–1.5 Zn–0.2 Ce, Mg–2.3 Zn–0.4 Mn–0.2 Ce, Mg–2 Gd–1 Zn, and Mg–3 Gd–1 Zn (weight %) alloys all satisfy the ductility condition, in agreement with observed weak texture, high ductility, and high formability (*13*, *34*, *38*, *39*). Mg-Zn-Zr alloys near the limits of Zr solubility are also predicted to reach χ > 2, in agreement with the high ductility seen in solution-treated Mg–0.5 Zn–0.6 Zr (*28*). Mg-Zn-Mn alloys at 1 wt % Mn show χ > 2 at low Zn, indicating that even dilute Mn may overcome the detrimental effects of Zn. Experiments show a range of ductilities, but they tend to saturate with increasing Zn, in spite of increasing basal strengthening and other mechanisms such as softened prism slip (*40*). The Mg–1.5 Zn–0.1 Ca (weight %) alloy has χ < 1, below the ductility condition, whereas further increase of Zn leads to less favorable conditions for ductility, consistent with experiments (*19*).

The theory is generally successful in rationalizing observed ductility across a wide range of existing ternary and higher alloys. This success of the theory, supported by the experimental validation of the physical mechanism, could enable quantitative refinements of a wide range of existing alloys, as well as guide searches for new Mg alloy compositions with high ductility. The theory does not address solubility or precipitation, which limit attainable solid-solution compositions and can impart strength and reduce ductility. Thermomechanical processing for favorable textures also remains a critical aspect of attaining improved ductility. Nonetheless, the theory provides mechanistic insight into which solutes and concentrations could enhance ductility and can contribute to accelerating the development of high-ductility, high-formability Mg alloys.

REFERENCES AND NOTES

1. Z. Wu, W. A. Curtin, *Nature* **526**, 62–67 (2015).
2. S. R. Agnew, J. W. Senn, J. A. Horton, *JOM* **58**, 62–69 (2006).
3. S. Sandlöbes, S. Zaefferer, I. Schestakow, S. Yi, R. Gonzalez-Martinez, *Acta Mater.* **59**, 429–439 (2011).
4. S. Sandlöbes *et al*., *Acta Mater.* **70**, 92–104 (2014).
5. B. Yin, Z. Wu, W. A. Curtin, *Acta Mater.* **136**, 249–261 (2017).
6. S. R. Agnew, M. H. Yoo, C. N. Tomé, *Acta Mater.* **49**, 4277–4289 (2001).
7. J. P. Hadorn *et al*., *Metall. Mater. Trans. A Phys. Metall. Mater. Sci.* **43**, 1347–1362 (2012).
8. A. Imandoust *et al*., *Acta Mater.* **138**, 27–41 (2017).
9. Z. Wu, W. A. Curtin, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 11137–11142 (2016).
10. Materials and methods are available as supplementary materials.
11. K. Hantzsche *et al*., *Scr. Mater.* **63**, 725–730 (2010).
12. R. K. Mishra *et al*., *Scr. Mater.* **59**, 562–565 (2008).
13. Y. Chino, X. Huang, K. Suzuki, K. Sassa, M. Mabuchi, *Mater. Sci. Eng. A* **528**, 566–572 (2010).
14. L. Yang *et al*., *Acta Biomater.* **9**, 8499–8508 (2013).
15. M. Qian, A. Das, *Scr. Mater.* **54**, 881–886 (2006).
16. R. Arroyave, D. Shin, Z. K. Liu, *Calphad* **29**, 230–238 (2005).
17. S.-Y. Chang, H. Tezuka, A. Kamio, *Mater. Trans. JIM* **38**, 526–535 (1997).
18. Y. Chino, M. Kado, T. Ueda, M. Mabuchi, *Metall. Mater. Trans. A Phys. Metall. Mater. Sci.* **42**, 1965–1973 (2011).
19. Y. Chino *et al*., *Mater. Trans.* **52**, 1477–1482 (2011).
20. H. Somekawa, A. Kinoshita, A. Kato, *Mater. Sci. Eng. A* **697**, 217–223 (2017).
21. C. H. Cáceres, D. M. Rovera, *J. Light Met.* **1**, 151–156 (2001).
22. S. Sandlöbes, M. Friák, J. Neugebauer, D. Raabe, *Mater. Sci. Eng. A* **576**, 61–68 (2013).
23. S. Ando, A. Kodera, K. Fukushima, M. Tsushida, H. Kitahara, *Mater. Sci. Forum* **783-786**, 341–345 (2014).
24. Y. Chino, M. Kado, M. Mabuchi, *Acta Mater.* **56**, 387–394 (2008).
25. X. Huang, K. Suzuki, Y. Chino, M. Mabuchi, *Mater. Sci. Eng. A* **633**, 144–153 (2015).
26. T. Nakata *et al*., *Mater. Sci. Eng. A* **673**, 443–449 (2016).
27. T. Nakata *et al*., *Acta Mater.* **130**, 261–270 (2017).
28. C. H. Cáceres, A. Blake, *Phys. Status Solidi* **194**, 147–158 (2002).
29. W. Tang, R. Chen, E. Han, *Mater. Trans.* **49**, 957–962 (2008).
30. E. Zhang, D. Yin, L. Xu, L. Yang, K. Yang, *Mater. Sci. Eng. C* **29**, 987–993 (2009).
31. A. A. Luo, R. K. Mishra, A. K. Sachdev, *Scr. Mater.* **64**, 410–413 (2011).
32. T. J. Chen, W. Wang, D. H. Zhang, Y. Ma, Y. Hao, *Mater. Des.* **44**, 555–565 (2013).
33. Y. Chino, M. Kado, M. Mabuchi, *Mater. Sci. Eng. A* **494**, 343–349 (2008).
34. Y. Chino, K. Sassa, M. Mabuchi, *Mater. Trans.* **49**, 2916–2918 (2008).
35. Y. Chino, K. Sassa, M. Mabuchi, *Mater. Sci. Eng. A* **513–514**, 394–400 (2009).
36. Y. Chino, K. Sassa, X. Huang, K. Suzuki, M. Mabuchi, *J. Jpn. Inst. Met. Mater.* **75**, 35–41 (2011).
37. X. Huang, K. Suzuki, Y. Chino, M. Mabuchi, *J. Alloys Compd.* **632**, 94–102 (2015).
38. L. Gao, H. Yan, J. Luo, A. A. Luo, R. Chen, *J. Magnesium Alloys* **1**, 283–291 (2013).
39. D. Wu, R. S. Chen, E. H. Han, *J. Alloys Compd.* **509**, 2856–2863 (2011).
40. A. Akhtar, E. Teghtsoonian, *Acta Metall.* **17**, 1351–1356 (1969).

ACKNOWLEDGMENTS

The authors acknowledge financial support of this work through a grant from the Swiss National Science Foundation entitled "Control of atomistic mechanisms of flow in magnesium alloys to achieve high ductility" (project #162350). The authors also acknowledge support from EPFL to the Laboratory for Multiscale Mechanics Modeling that enabled the required high-performance computing provided by Scientific and IT Application Support (SCITAS) at EPFL. All data are reported in the main paper and supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6374/447/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S7
Table S1
References (*41*–*150*)

5 September 2017; accepted 20 December 2017
10.1126/science.aap8716

Cite as: P. A. Murthy *et al.*, *Science* 10.1126/science.aan5950 (2017).

High-temperature pairing in a strongly interacting two-dimensional Fermi gas

Puneet A. Murthy,[1]*† Mathias Neidig,[1]† Ralf Klemt,[1]† Luca Bayha,[1] Igor Boettcher,[2] Tilman Enss,[3] Marvin Holten,[1] Gerhard Zürn,[1] Philipp M. Preiss,[1] Selim Jochim[1]

[1]Physics Institute, Heidelberg University, Heidelberg, Germany. [2]Department of Physics, Simon Fraser University, Burnaby, BC, Canada. [3]Institute for Theoretical Physics, Heidelberg University, Heidelberg, Germany.

*Corresponding author. Email: murthy@physi.uni-heidelberg.de

†These authors contributed equally to this work.

The nature of the normal phase of strongly correlated fermionic systems is an outstanding question in quantum many-body physics. We use spatially resolved radio-frequency spectroscopy to measure pairing energy of fermions across a wide range of temperatures and interaction strengths in a two-dimensional gas of ultracold fermionic atoms. We observe many-body pairing at temperatures far above the critical temperature for superfluidity. In the strongly interacting regime, the pairing energy in the normal phase significantly exceeds the intrinsic two-body binding energy of the system and shows a clear dependence on local density. This implies that pairing in this regime is driven by many-body correlations, rather than two-body physics. Our findings show that pairing correlations in strongly interacting two-dimensional fermionic systems are remarkably robust against thermal fluctuations.

Fermion pairing is the key ingredient for superconductivity and superfluidity in fermionic systems (*1*). In a system with s-wave interactions, two scenarios can occur: In the first one, as realized for weakly attractive fermions that are described by the theory of Bardeen-Cooper-Schrieffer (BCS), formation and condensation of pairs both take place at the same critical temperature (T_c) (*2*). In the second scenario, fermion pairing accompanied by a suppression of the density of states at the Fermi surface occurs at temperatures exceeding the critical temperature. Finding a description of this so-called 'pseudogap' phase, especially for two-dimensional (2D) systems, is thought to be a promising route to understanding the complex physics of high temperature superconductivity (*3*–*6*).

The Bose-Einstein Condensation (BEC)–BCS crossover of ultracold atoms constitutes a versatile framework to explore the normal phase of strongly correlated fermions (Fig. 1A). The crossover smoothly connects two distinct regimes of pairing: the BEC regime of tightly bound molecules and the BCS regime of weakly bound Cooper pairs. In 2D (unlike 3D) systems with contact interactions, a two-body bound state with binding energy E_B exists for arbitrarily small attractions between the atoms. The interactions in the many-body system are captured by the dimensionless parameter $\ln(k_F a_{2D})$ where k_F is the Fermi momentum and a_{2D} is the 2D scattering length. As we tune the interaction strength from the BEC [large negative $\ln(k_F a_{2D})$] to the BCS side [large positive $\ln(k_F a_{2D})$], the character of the system smoothly changes from bosonic to fermionic (*7*). A strongly interacting region lies in between these two weakly interacting limits where a_{2D} is of the same order as the inter-particle spacing ($\sim k_F^{-1}$). Previously, a matter-wave focusing method was used to measure the pair momentum distribution of a 2D Fermi gas across the crossover, leading to the observation of the Berezinskii-Kosterlitz-Thouless (BKT) transition at low temperatures (*8*, *9*). An outstanding question concerns the nature of the normal phase above the critical temperature—specifically, how does the normal phase cross over from a gapless Fermi liquid on the weakly interacting BCS side to a Bose liquid of two-body dimers on the BEC side? Is there an interaction regime where pairing is driven by many-body effects rather than the two-body bound state? Although previous cold atom experiments have explored this problem both in 3D (*10*–*14*) and 2D (*15*, *16*) systems, a consensus is yet to emerge (*3*, *7*, *17*–*20*).

Here, we address these questions by studying the normal phase of such a 2D ultracold Fermi gas trapped in a harmonic potential. The underlying potential leads to an inhomogeneous density distribution, and therefore we can use the local density approximation to directly measure the density dependence of many-body properties. We perform our experiments with a two-component mixture of ^{6}Li atoms with approximately 3×10^4 particles per spin state that are loaded into a single layer of an anisotropic harmonic optical trap. The trap frequencies are $\omega_z \approx 2\pi \times 6.95$ kHz and $\omega_r \approx 2\pi \times 22$ Hz in the axial and radial directions, leading to an aspect ratio of about 300:1. We reach the kinematic 2D regime by ensuring that the thermodynamic energy scales, temperature

(T) and chemical potential (μ), are smaller than the axial confinement energy (*21*, *22*). We tune the scattering length a_{2D} by means of a broad magnetic Feshbach resonance (*23*).

To investigate fermion pairing in our system, we use radio-frequency (RF) spectroscopy. We perform experiments with the three lowest lying hyperfine states of ^{6}Li which at low magnetic fields are given by: $|1\rangle = \left|F = \frac{1}{2}, m_F = -\frac{1}{2}\right\rangle$, $|2\rangle = \left|\frac{1}{2}, \frac{1}{2}\right\rangle$, and $|3\rangle = \left|\frac{3}{2}, -\frac{3}{2}\right\rangle$. We start with a two-component mixture of atoms in hyperfine states $|a\rangle|b\rangle \equiv |1\rangle|2\rangle$ or $|a\rangle|b\rangle \equiv |1\rangle|3\rangle$ (fig. S1) (*21*). An RF pulse transfers atoms from state $|b\rangle$ to the third unoccupied hyperfine state $|c\rangle$, and we subsequently image the remaining density distribution in $|b\rangle$. The idea underlying this technique is that the atomic transition frequencies between hyperfine states are shifted by interactions or pairing effects in an ensemble. For example, a state of coexisting pairs and free atoms (Fig. 1B) will lead to two energetically separated branches in the RF spectrum from which we can gain quantitative information on pairing and correlations in the many-body system. Creating initial samples in either $|1\rangle|2\rangle$ or $|1\rangle|3\rangle$ allows us to access a wide range of interaction strengths and minimize final state interaction effects (*21*).

In our inhomogeneous 2D system, the local Fermi energy depends on the local density $n(r)$ in each spin state according to $E_F(r) = (2\pi\hbar/m)n(r)$, where m is the mass of a ^{6}Li atom (*24*). As a consequence, the thermodynamic quantities T/T_F and $\ln(k_F a_{2D})$ also vary spatially across the cloud. We apply the thermometry developed in (*21*, *25*) to extract these local observables. We measure the local spectral response (*26*) by choosing a RF pulse duration (τ_{RF} = 4 ms) that is sufficiently short to prevent significant diffusion of transferred atoms, but also sufficiently long that we obtain an adequate Fourier limited frequency resolution $\delta\omega_{RF} \approx 2\pi \times 220$ Hz (fig. S2) (*21*). In Fig. 1C, we show a typical absorption image of the 2D cloud which is used as a reference and another with a RF pulse applied at a particular frequency. The difference between the two images features a spatial ring structure, which qualitatively shows that for a given frequency, the depletion of atoms in initial state $|b\rangle$ occurs at a well-defined density/radius. By performing this measurement for a range of RF frequencies, we can tomographically reconstruct the spatially resolved spectral response function

$I(r, \omega_{RF}) = [n_0(r) - n'(r, \omega_{RF})]/n_0(r)$

where $n_0(r)$ and $n'(r, \omega_{RF})$ are the density distribution of atoms in state $|b\rangle$ without and with the RF pulse. An example of the tomographically reconstructed spectra, taken at $\ln(k_F a_{2D}) \sim 1.5$, is shown in Fig. 1D. The frequency of maximum response depends smoothly on the radius and therefore the local density. Such density dependent shifts may arise from pairing effects - wherein the effective binding energy between fermions is dependent on the density of the medium, or Hartree shifts which are offsets in the spectrum caused by the mean-field interaction energy with no influence on the binding energy between fermions. The position of the RF absorption peak alone (Fig. 1D) does not serve as a reliable observable to distinguish between these two effects as it lacks a suitable reference energy that already incorporates Hartree shifts (*21*). One way to obtain this reference scale is to measure the RF transitions from both bound and free branches to the third unoccupied state (*21*). However, we find that in the temperature regime ($T/T_F < 1.5$) explored in our experiments, the thermal occupation of the free (unpaired) branch is too low to be observed.

In order to achieve a sufficient population of the unpaired branch, we apply the quasiparticle spectroscopy method pioneered in (*27*) for the measurement of the superfluid gap of a 3D Fermi gas. Although our system is in the normal phase, the same technique can be used to determine the pairing gap. The key idea of this method lies in creating a slightly spin-imbalanced mixture so that the excess majority atoms necessarily remain unpaired owing to the density mismatch. These unpaired atoms (or dressed quasiparticles) contribute a second absorption maximum in the RF response function in addition to the one from pairs. We refer to the energy difference between the two branches in the spectrum as the pairing energy ΔE. In our experiments, we create a slight spin-imbalance $P = \left(n_{|b\rangle} - n_{|a\rangle}\right)/\left(n_{|b\rangle} + n_{|a\rangle}\right) \lesssim 0.15$ using a sequence of Landau-Zener sweeps (*21*), where $n_{|b\rangle}$ and $n_{|a\rangle}$ are densities in hyperfine states $|a\rangle$ and $|b\rangle$. We show typical density profiles of majority and minority components in Fig. 2A.

The pairing energy ΔE allows us to distinguish between two different pairing scenarios. If ΔE coincides with the energy E_B of the dimer state, we are in the two-body regime. In contrast, we associate the situation of a density (E_F)-dependent ΔE, exceeding E_B, with many-body pairing wherein the relative pair wavefunction is strongly altered by the presence of the surrounding medium of interacting fermions. In Fig. 2, B and C, we illustrate these two scenarios using ideal single-particle dispersion relations in the BEC and BCS limits at zero temperature; both limits have free and bound branches. The RF photons drive transitions from these branches to the continuum. The transition of bound pairs occurs with a sharp onset at a threshold RF frequency where the dissociated fragments have no relative momenta. Higher frequency RF photons provide relative momenta to the transferred particles which leads to a slowly decaying tail in the

spectrum (*21*). This leads to the highly asymmetric feature seen in the spectrum in Fig. 2, D and E. On the other hand, the transition of unpaired particles leads to a symmetric peak as it does not involve a dissociation process.

The crucial difference between the BEC and BCS regimes arises from the fact that the energy minimum of the free branch occurs at $k \sim 0$ on the BEC side and $k \sim k_F$ on the BCS side. Although the RF spectra in the two limits seem qualitatively similar, the fundamental difference in their dispersion appears as an energy difference between the two branches. Whereas on the BEC side $\triangle E \sim E_B$ independent of local density; $\triangle E \sim \triangle + E_B$ on the BCS side, where $\Delta = \sqrt{2E_B E_F}$ is the many-body gap parameter from BCS theory. In the latter case, $\triangle E$ is necessarily larger than E_B and density-dependent. Although this idealized picture provides some intuition for RF spectroscopy in a 2D Fermi gas, the actual dispersion relations in the strongly interacting region and at high temperatures may not follow this mean-field description. However, the behavior of $\triangle E$—particularly its deviation from E_B—is still a reliable indicator for pairing beyond two-body physics.

In Fig. 3, A and B, we show the measured spectra $I(r, \omega)$ for magnetic fields 670 G and 690 G using a $|1\rangle|3\rangle$ mixture, which corresponds to central values of $\ln(k_F a_{2D}) \sim -0.5$ and $\ln(k_F a_{2D}) \sim 1$, respectively. The response from unpaired quasiparticles appears at frequency $\omega_{RF} \sim 0$, whereas the pairing branch with an asymmetric line shape appears at larger frequencies. Examples of spectra at fixed radii are shown in Fig. 3, C and D. We fit these local spectra with a combined fit function that includes a symmetric Gaussian (for the quasiparticle peak) and an asymmetric threshold function (for the paired peak) that is convolved with a Gaussian to account for spectral broadening arising from finite RF frequency resolution and final state effects (*28*). We present a detailed account of our data analysis in (*21*). The choice of fit function has a systematic effect on the quantitative results presented here, which cannot be eliminated at this point because a reliable theoretical prediction of the shape of the spectral function exists only in the weakly coupled BEC (*18*) and BCS (*29*) limits.

At a qualitative level, the main observations from Fig. 3 are the following. Both branches in the spectra show density dependence, part of which can be attributed to a Hartree shift. Adding the binding energy E_B to the quasi-particle branch yields the two-body expectation for the threshold position. This picture is applicable to the whole spectrum in Fig. 3A, which corresponds to a measurement on the BEC side of the crossover. In contrast, for the spectrum displayed in Fig. 3B, corresponding to the crossover regime, we observe $\triangle E \sim E_B$ only in the outer regions of the cloud where the density is low enough that only the two-body bound state plays a role. Toward the center of the cloud, $\triangle E$ begins to significantly exceed E_B and shows a strong dependence on the local density (E_F), indicating that pairing in this regime is a many-body phenomenon. At very low temperatures, the measurement of $\triangle E$ is difficult because the occupation of the free branch is too low, and for this work we were unable to prepare a spin-imbalanced sample at temperatures below $T/T_F \sim 0.4$ (*21*). However, for a balanced gas, we qualitatively observe that the threshold position of the bound branch increases continuously with decreasing temperature, even as we cross the superfluid transition. This indicates that in the crossover regime, a many-body gap opens in the normal phase rather than at $T_c \approx 0.17 T_F$ as expected from BCS theory (fig. S8) (*21*). This observation is the first main result of this work.

To quantitatively study the change in the nature of pairing from the BEC to the BCS side, we measure the spectra at different magnetic fields and extract $\triangle E$ in units of the two-body binding energy E_B. In Fig. 4A, we plot the temperature dependence of $\triangle E/E_B$ for different interaction strengths, and Fig. 4B shows the variation of $\triangle E/E_B$ as a function of $\ln(k_F a_{2D})$ for a fixed ratio $T/T_F \approx 0.5$. This constitutes an extremely high temperature regime even in the context of ultracold fermionic superfluidity, where the largest observed critical temperatures are $T_c/T_F \approx 0.17$ (*8*, *9*). We perform our measurements with both $|1\rangle|2\rangle$ and $|1\rangle|3\rangle$ mixtures (blue and red points in Fig. 4B) in an overlapping interaction regime. The two mixtures differ significantly in their final state interaction strengths, yet they show similar values of $\triangle E$ around $\ln(k_F a_{2D}) \approx 0.5$ demonstrating the robustness of the quantity $\triangle E$ against final state effects. For larger $\ln(k_F a_{2D})$, the two mixtures allow us to probe complementary regions of the crossover. Details of the experimental parameters used for the two mixtures are tabulated in (table S1) (*21*).

In Fig. 4, we observe that for $\ln(k_F a_{2D}) \leq 0.5$ the spectra are well-described by two-body physics. In contrast, the pronounced density-dependent gap significantly exceeding E_B for $\ln(k_F a_{2D}) \geq 0.5$ signals the crossover to a many-body pairing regime. In particular, we observe that $\triangle E/E_B$ peaks at $\ln(k_F a_{2D}) \sim 1$, where $\triangle E \sim 2.6 E_B$ and is a significant fraction of $E_F (0.6 E_F)$. The identification of this strongly correlated many-body pairing regime and the observation of many-body induced pairing at temperatures several times the critical temperature is the second main result of this work. For larger $\ln(k_F a_{2D})$, we see a downward trend in $\triangle E/E_B$, and for $\ln(k_F a_{2D}) > 1.5$, we observe only a single branch in the spectra near $\omega_{RF} \sim 0$, suggesting the absence of a gap larger than the scale of our experimental resolution (fig. S6) (*21*). Our qualitative observation of a vanishing gap for weaker attraction is consistent with the picture of the normal phase in the BCS limit being a gapless Fermi liquid (*30*). Interestingly, the non-monotonous behavior of $\triangle E$ as a function of $\ln(k_F a_{2D})$, as shown

in Fig. 4B, is also qualitatively predicted by finite-temperature BCS theory (fig. S4) (*21*) for the superfluid phase.

We now turn to a discussion of our results in the context of current theoretical understanding and previous experimental work. In (*15*), Sommer *et al.* performed trap-averaged RF spectroscopy in the 3D-2D crossover and found good agreement with the mean-field two-body expectation in the regime $\ln(k_F a_{2D}) \leq 0.5$. In (*16*), Feld *et al.* observed signatures of pairing in the normal phase using momentum-resolved (but trap-averaged) spectroscopy, in a similar interaction regime as (*15*) which were interpreted as a many-body pseudogap. However, subsequent theoretical work based on two-body physics only (*18*, *19*) was consistent with that many of the observations in (*16*). Beyond this previously explored regime, our measurements reveal that many-body effects enhance the pairing energy far above the critical temperature, with the maximum enhancement occurring at $\ln(k_F a_{2D}) \approx 1$, where a reliable mean-field description is not available. With regard to the long-standing question concerning the nature of the normal phase of a strongly interacting Fermi gas (*7*, *17*, *31*–*33*), our experiments reveal the existence of a state in the phase diagram whose behavior deviates from both Bose Liquid and Fermi liquid descriptions. Finding a complete description of this strongly correlated phase is an exciting challenge for both theory and experiment.

REFERENCES AND NOTES

1. C. N. Yang, Concept of off-diagonal long-range order and the quantum phases of liquid He and of superconductors. *Rev. Mod. Phys.* **34**, 694–704 (1962). doi:10.1103/RevModPhys.34.694
2. J. Bardeen, L. N. Cooper, J. R. Schrieffer, Theory of superconductivity. *Phys. Rev.* **108**, 1175–1204 (1957). doi:10.1103/PhysRev.108.1175
3. E. J. Mueller, Review of pseudogaps in strongly interacting Fermi gases. *Rep. Prog. Phys.* **80**, 104401 (2017). doi:10.1088/1361-6633/aa7e53 Medline
4. Q. Chen, Y. He, C.-C. Chien, K. Levin, Theory of radio-frequency spectroscopy experiments in ultracold Fermi gases and their relation to photoemission cuprates. *Rep. Prog. Phys.* **72**, 122501 (2009). doi:10.1088/0034-4885/72/12/122501
5. J. J. Lee, F. T. Schmitt, R. G. Moore, S. Johnston, Y.-T. Cui, W. Li, M. Yi, Z. K. Liu, M. Hashimoto, Y. Zhang, D. H. Lu, T. P. Devereaux, D.-H. Lee, Z.-X. Shen, Interfacial mode coupling as the origin of the enhancement of T(c) in FeSe films on SrTiO3. *Nature* **515**, 245–248 (2014). doi:10.1038/nature13894 Medline
6. S. Kasahara, T. Yamashita, A. Shi, R. Kobayashi, Y. Shimoyama, T. Watashige, K. Ishida, T. Terashima, T. Wolf, F. Hardy, C. Meingast, H. V. Löhneysen, A. Levchenko, T. Shibauchi, Y. Matsuda, Giant superconducting fluctuations in the compensated semimetal FeSe at the BCS-BEC crossover. *Nat. Commun.* **7**, 12843 (2016). doi:10.1038/ncomms12843 Medline
7. J. Levinsen, M. M. Parish, Strongly interacting two-dimensional Fermi gases. *Ann. Rev. Cold Atom. Mol.* **3**, 1–75 (2015). doi:10.1142/9789814667746_0001
8. M. G. Ries, A. N. Wenz, G. Zürn, L. Bayha, I. Boettcher, D. Kedar, P. A. Murthy, M. Neidig, T. Lompe, S. Jochim, Observation of pair condensation in the quasi-2D BEC-BCS Crossover. *Phys. Rev. Lett.* **114**, 230401 (2015). doi:10.1103/PhysRevLett.114.230401 Medline
9. P. A. Murthy, I. Boettcher, L. Bayha, M. Holzmann, D. Kedar, M. Neidig, M. G. Ries, A. N. Wenz, G. Zürn, S. Jochim, Observation of the Berezinskii-Kosterlitz-Thouless phase transition in an ultracold Fermi gas. *Phys. Rev. Lett.* **115**, 010401 (2015). doi:10.1103/PhysRevLett.115.010401 Medline
10. S. Nascimbène, N. Navon, S. Pilati, F. Chevy, S. Giorgini, A. Georges, C. Salomon, Fermi-liquid behavior of the normal phase of a strongly interacting gas of cold atoms. *Phys. Rev. Lett.* **106**, 215303 (2011). doi:10.1103/PhysRevLett.106.215303 Medline
11. J. P. Gaebler, J. T. Stewart, T. E. Drake, D. S. Jin, A. Perali, P. Pieri, G. C. Strinati, Observation of pseudogap behaviour in a strongly interacting Fermi gas. *Nat. Phys.* **6**, 569–573 (2010). doi:10.1038/nphys1709
12. A. Perali, F. Palestini, P. Pieri, G. C. Strinati, J. T. Stewart, J. P. Gaebler, T. E. Drake, D. S. Jin, Evolution of the normal state of a strongly interacting Fermi gas from a pseudogap phase to a molecular Bose gas. *Phys. Rev. Lett.* **106**, 060402 (2011). doi:10.1103/PhysRevLett.106.060402 Medline
13. Y. Sagi, T. E. Drake, R. Paudel, R. Chapurin, D. S. Jin, Breakdown of the fermi liquid description for strongly interacting fermions. *Phys. Rev. Lett.* **114**, 075301 (2015). doi:10.1103/PhysRevLett.114.075301 Medline
14. C. H. Schunck, Y. Shin, A. Schirotzek, M. W. Zwierlein, W. Ketterle, Pairing without superfluidity: The ground state of an imbalanced Fermi mixture. *Science* **316**, 867–870 (2007). doi:10.1126/science.1140749 Medline
15. A. T. Sommer, L. W. Cheuk, M. J. H. Ku, W. S. Bakr, M. W. Zwierlein, Evolution of fermion pairing from three to two dimensions. *Phys. Rev. Lett.* **108**, 045302 (2012). doi:10.1103/PhysRevLett.108.045302 Medline
16. M. Feld, B. Fröhlich, E. Vogt, M. Koschorreck, M. Köhl, Observation of a pairing pseudogap in a two-dimensional Fermi gas. *Nature* **480**, 75–78 (2011). doi:10.1038/nature10627 Medline
17. P. Törmä, Physics of ultracold Fermi gases revealed by spectroscopies. *Phys. Scr.* **91**, 043006 (2016). doi:10.1088/0031-8949/91/4/043006
18. M. Barth, J. Hofmann, Pairing effects in the nondegenerate limit of the two-dimensional Fermi gas. *Phys. Rev. A* **89**, 013614 (2014). doi:10.1103/PhysRevA.89.013614
19. V. Ngampruetikorn, J. Levinsen, M. M. Parish, Pair correlations in the two-dimensional Fermi gas. *Phys. Rev. Lett.* **111**, 265301 (2013). doi:10.1103/PhysRevLett.111.265301 Medline
20. F. Marsiglio, P. Pieri, A. Perali, F. Palestini, G. C. Strinati, Pairing effects in the normal phase of a two-dimensional Fermi gas. *Phys. Rev. B* **91**, 054509 (2015). doi:10.1103/PhysRevB.91.054509
21. Materials and methods are available as supplementary materials.
22. P. Dyke, K. Fenech, T. Peppler, M. G. Lingham, S. Hoinka, W. Zhang, S.-G. Peng, B. Mulkerin, H. Hu, X.-J. Liu, C. J. Vale, Criteria for two-dimensional kinematics in an interacting Fermi gas. *Phys. Rev. A* **93**, 011603 (2016). doi:10.1103/PhysRevA.93.011603
23. G. Zürn, T. Lompe, A. N. Wenz, S. Jochim, P. S. Julienne, J. M. Hutson, Precise characterization of 6Li Feshbach resonances using trap-sideband-resolved RF spectroscopy of weakly bound molecules. *Phys. Rev. Lett.* **110**, 135301 (2013). doi:10.1103/PhysRevLett.110.135301 Medline
24. E_F/h has typical values of 7.5 kHz at the center of the cloud.
25. I. Boettcher, L. Bayha, D. Kedar, P. A. Murthy, M. Neidig, M. G. Ries, A. N. Wenz, G. Zürn, S. Jochim, T. Enss, Equation of state of ultracold fermions in the 2D BEC-BCS Crossover region. *Phys. Rev. Lett.* **116**, 045303 (2016). doi:10.1103/PhysRevLett.116.045303 Medline
26. Y. Shin, C. H. Schunck, A. Schirotzek, W. Ketterle, Tomographic rf spectroscopy of a trapped Fermi gas at unitarity. *Phys. Rev. Lett.* **99**, 090403 (2007). doi:10.1103/PhysRevLett.99.090403 Medline
27. A. Schirotzek, Y. I. Shin, C. H. Schunck, W. Ketterle, Determination of the superfluid gap in atomic Fermi gases by quasiparticle spectroscopy. *Phys. Rev. Lett.* **101**, 140403 (2008). doi:10.1103/PhysRevLett.101.140403 Medline
28. C. Langmack, M. Barth, W. Zwerger, E. Braaten, Clock shift in a strongly interacting two-dimensional Fermi gas. *Phys. Rev. Lett.* **108**, 060402 (2012). doi:10.1103/PhysRevLett.108.060402 Medline
29. A. M. Fischer, M. M. Parish, Quasi-two-dimensional Fermi gases at finite temperatures. *Phys. Rev. B* **90**, 214503 (2014). doi:10.1103/PhysRevB.90.214503
30. B. Fröhlich, M. Feld, E. Vogt, M. Koschorreck, M. Köhl, C. Berthod, T. Giamarchi, Two-dimensional Fermi liquid with attractive interactions. *Phys. Rev. Lett.* **109**, 130403 (2012). doi:10.1103/PhysRevLett.109.130403 Medline

31. C. A. Sá de Melo, M. Randeria, J. R. Engelbrecht, Crossover from BCS to Bose superconductivity: Transition temperature and time-dependent Ginzburg-Landau theory. *Phys. Rev. Lett.* **71**, 3202–3205 (1993). doi:10.1103/PhysRevLett.71.3202 Medline
32. M. Randeria, J.-M. Duan, L.-Y. Shieh, Cooper pairs and composite Bosons in two dimensions. *Physica C* **162-164**, 1457–1458 (1989). doi:10.1016/0921-4534(89)90770-3
33. M. Randeria, J.-M. Duan, L.-Y. Shieh, Superconductivity in a two-dimensional Fermi gas: Evolution from Cooper pairing to Bose condensation. *Phys. Rev. B Condens. Matter* **41**, 327–343 (1990). doi:10.1103/PhysRevB.41.327 Medline
34. D. Mitra, P. T. Brown, P. Schauß, S. S. Kondov, W. S. Bakr, Phase separation and pair condensation in a spin-imbalanced 2D Fermi gas. *Phys. Rev. Lett.* **117**, 093601 (2016). doi:10.1103/PhysRevLett.117.093601 Medline
35. M. Koschorreck, D. Pertot, E. Vogt, B. Fröhlich, M. Feld, M. Köhl, Attractive and repulsive Fermi polarons in two dimensions. *Nature* **485**, 619–622 (2012). doi:10.1038/nature11151 Medline

ACKNOWLEDGMENTS

We gratefully acknowledge insightful discussions with Meera Parish, Jesper Levinsen, Nicolò Defenu and Wilhelm Zwerger. We would like to thank Thomas Lompe for discussions and for a critical reading of the manuscript. This work has been supported by the ERC consolidator grant 725636, the Heidelberg Center for Quantum Dynamics and is part of the DFG Collaborative Research Centre "SFB 1225 (ISOQUANT)". IB acknowledges support from DFG und grant no. BO 4640/1-1. P.M.P. acknowledges funding from European Union's Horizon 2020 program under the Marie Sklodowska-Curie grant agreement No. 706487. Supporting data can be found in the Supplementary Material. Raw data are available upon request. P.A.M and G.Z. initiated the project. P.A.M., M.N., R.K. and M.H. performed the measurements and analyzed the data. I.B. and T.E. provided theory support and assistance with preparing the manuscript. P.M.P. and S.J. supervised the project. All authors contributed to the interpretation and discussion of the experimental results.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/cgi/content/full/science.aan5950/DC1
Materials and Methods
Supplementary text
Fig. S1 to S8
Table S1
References (*34*, *35*)

5 May 2017; accepted 6 December 2017
Published online 21 December 2017
10.1126/science.aan5950

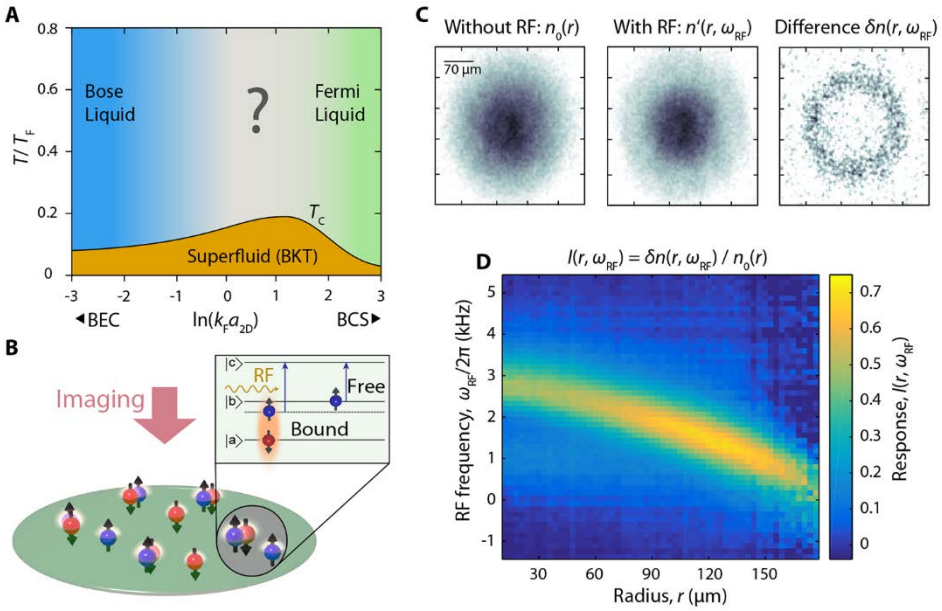


Fig. 1. Exploring fermion pairing in a strongly interacting 2D Fermi gas. (**A**) Schematic phase diagram of the BEC – BCS crossover. In this work we investigate the nature of pairing in the normal phase of the crossover regime between the weakly interacting Bose and Fermi liquids. (**B**) Illustration of radio-frequency (RF) spectroscopy of a 2D two-component Fermi gas. Pairing and many-body effects shift the atomic transition frequencies between the hyperfine states $|b\rangle - |c\rangle$, which results in observable signatures in the RF response of the system. (**C**) Absorption images of the cloud [taken at $\ln(k_F a_{2D}) \approx 1.5$ and $T/T_F \sim 0.3$] without RF (reference) and with RF at a particular frequency, and the difference between the two images. The ring feature in $\delta n(r)$ reveals the density dependence of the RF response. (**D**) Spatially resolved spectral response function reconstructed from absorption images taken at different RF frequencies. At low temperatures in the spin-balanced sample, the occupation of the free particle branch is too low to be observable, which makes it difficult to distinguish between mean-field shifts and pairing effects.

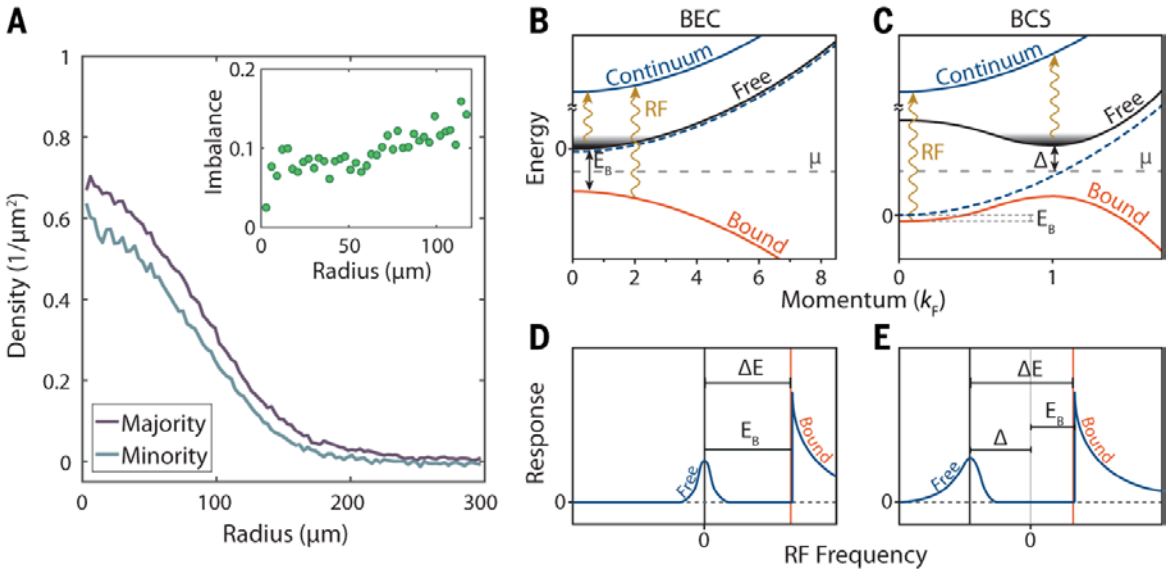


Fig. 2. Quasiparticle spectroscopy in the BEC and BCS limits. (**A**) We create a slightly imbalanced mixture of hyperfine states to artificially populate the free particle branch. The density distributions of the majority and minority spins are shown, as well as the corresponding local imbalance (inset). (**B** and **C**) Schematic illustration of single particle dispersion relations in the BEC and BCS limits at zero temperature. Paired atoms reside in the lowest branch (bound), and are transferred to the continuum of unoccupied states. The excess majority atoms are unpaired and occupy the upper quasiparticle (free) branch in the spectrum preferentially at $k \sim 0$ (BEC) and $k \sim k_F$ (BCS). The energy difference between the free-particle dispersion in a non-interacting system (blue dashed line) and the continuum (blue solid line) is the bare hyperfine transition energy and serves as the reference for (D) and (E). (**D** and **E**) The transition of paired atoms into the continuum yields an asymmetric response with a sharp threshold in the RF spectral function. The quasi-particle transition contributes another peak which appears at $\omega_{RF} = 0$ on the BEC side and $\omega_{RF} = -\Delta$ on the BCS side, where Δ is the BCS gap parameter. Their relative difference yields the pairing energy ΔE which reveals the distinction between two-body ($\Delta E \sim E_B$) and many-body pairing ($\Delta E > E_B$) in the two limits.

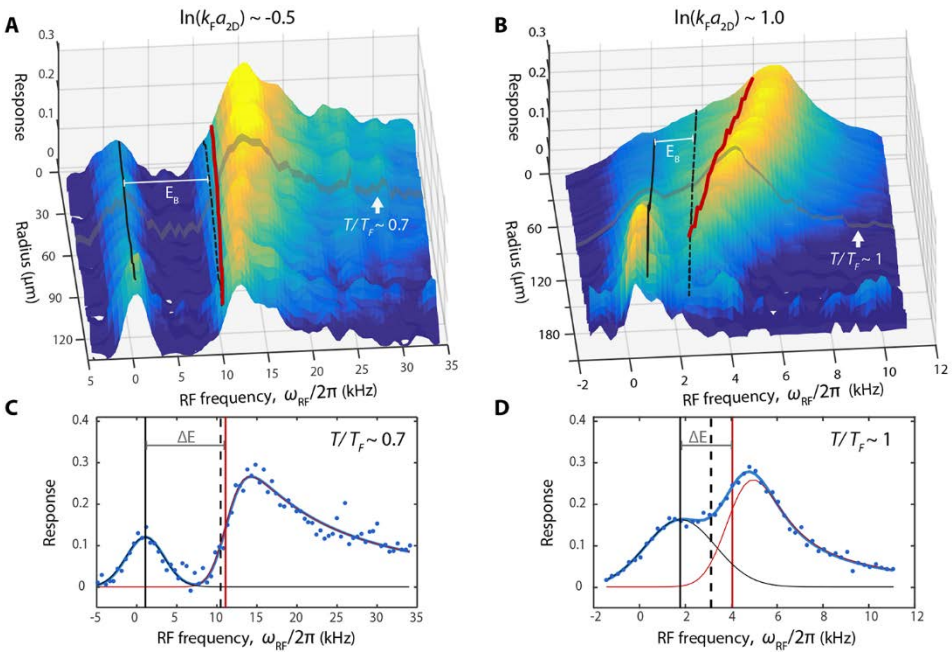


Fig. 3. From two-body dimers to many-body pairing. The spatially resolved response function $I(r, \omega_{RF})$ shows qualitatively different behavior for two different scattering lengths. (**A** and **B**) $I(r, \omega_{RF})$ for central ln($k_F a_{2D}$) ~ −0.5 and 1.0, respectively. The gray lines correspond to local T/T_F ~ 0.7 in (A) and T/T_F ~ 1 in (B). The 3D visualization is obtained using a linear interpolation between 3000 data points each of which is an average of 30 realizations. The black solid line is the peak position of the free branch, the red line is the threshold position of the bound branch and the black dashed line is displaced from the free peak by the two-body binding energy E_B. The energy difference between free and bound branches is the pairing energy ΔE, which is seen to agree with E_B in (A) (BEC regime), but significantly exceeds E_B in (B) (crossover regime). The differential density-dependence of the energy of the two branches implies that the pair wavefunction is strongly modified by the many-body system. (**C** and **D**) Local spectra at a fixed radius indicate by gray lines in (A) and (B) corresponding to a homogeneous system with T/T_F ~ 0.7 and 1 respectively. The solid blue curves are the fits to the data; the black and red curves are Gaussian and threshold fits to the two branches (*21*).

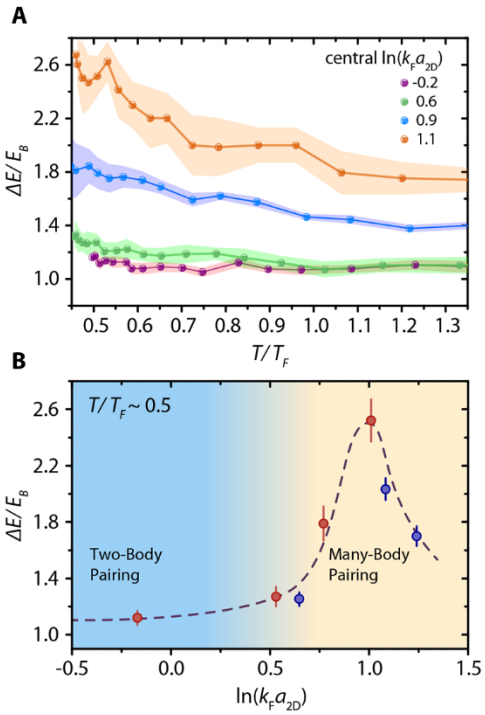


Fig. 4. Normal phase in the 2D BEC-BCS crossover regime. (A) Pairing energy ΔE in units of E_B plotted as a function of T/T_F for different interaction strengths [central ln($k_F a_{2D}$)]. Each point in (A) is the result of fits to local spectra (Fig. 3, C and D) which are averaged over 30 shots (**B**) Many-body induced high temperature pairing. We plot $\Delta E/E_B$ as a function of ln($k_F a_{2D}$) for fixed ratio T/T_F ~ 0.5. Red and blue circles correspond to measurements taken with $|1\rangle|3\rangle$ and $|1\rangle|2\rangle$ mixtures. The dashed black line is a guide to the eye. The errors indicated as shaded bands in (A) and bars in (B) are obtained from the fitting procedure explained in (*21*). For ln($k_F a_{2D}$) ≤ 0.5 (strong attraction) we have $\Delta E/E_B \sim 1$ with negligible density-dependence, indicating two-body pairing. For larger ln($k_F a_{2D}$) (less attraction), $\Delta E/E_B$ significantly exceeds 1 and reaches a maximum of 2.6 before showing a downward trend. At ln($k_F a_{2D}$) ~ 1, we have a critical temperature of $T_c \sim 0.17 T_F$ (*8*), which indicates the onset of many-body pairing at temperatures several times T_c.

PALEOANTHROPOLOGY

The earliest modern humans outside Africa

Israel Hershkovitz,[1,2]*† Gerhard W. Weber,[3]† Rolf Quam,[4,5,6]† Mathieu Duval,[7,8] Rainer Grün,[7,9] Leslie Kinsley,[9] Avner Ayalon,[10] Miryam Bar-Matthews,[10] Helene Valladas,[11] Norbert Mercier,[12] Juan Luis Arsuaga,[5,13] María Martinón-Torres,[8,14] José María Bermúdez de Castro,[8,14] Cinzia Fornai,[3,15] Laura Martín-Francés,[8,16] Rachel Sarig,[2,17] Hila May,[1,2] Viktoria A. Krenn,[3,15] Viviane Slon,[1] Laura Rodríguez,[5,18,19] Rebeca García,[5,18] Carlos Lorenzo,[20,21] Jose Miguel Carretero,[5,18] Amos Frumkin,[22] Ruth Shahack-Gross,[23] Daniella E. Bar-Yosef Mayer,[24,25] Yaming Cui,[26] Xinzhi Wu,[26] Natan Peled,[27] Iris Groman-Yaroslavski,[28] Lior Weissbrod,[28] Reuven Yeshurun,[28] Alexander Tsatskin,[28] Yossi Zaidner,[28,29] Mina Weinstein-Evron[28]

To date, the earliest modern human fossils found outside of Africa are dated to around 90,000 to 120,000 years ago at the Levantine sites of Skhul and Qafzeh. A maxilla and associated dentition recently discovered at Misliya Cave, Israel, was dated to 177,000 to 194,000 years ago, suggesting that members of the *Homo sapiens* clade left Africa earlier than previously thought. This finding changes our view on modern human dispersal and is consistent with recent genetic studies, which have posited the possibility of an earlier dispersal of *Homo sapiens* around 220,000 years ago. The Misliya maxilla is associated with full-fledged Levallois technology in the Levant, suggesting that the emergence of this technology is linked to the appearance of *Homo sapiens* in the region, as has been documented in Africa.

The timing and routes of modern human migration out of Africa are key issues for understanding the evolution of our own species. The fossil evidence suggests that the earliest members of the *Homo sapiens* clade (Jebel Irhoud, Omo, and Herto) appeared in Africa during the late Middle Pleistocene (*1*–*4*). Outside Africa, modern humans appeared much later, during the Late Pleistocene in the Levant (Qafzeh, Skhul) (*5*–*7*), and possibly in East Asia (Daoxian) (*8*). Misliya Cave, Israel, is part of a complex of prehistoric caves along the western slopes of Mount Carmel (Fig. 1 and fig. S1). Here we report on an adult hominin left hemimaxilla (Misliya-1) (Fig. 2A) found in Square N9 of the upper part of the Early Middle Palaeolithic (EMP) archaeological layer of the site (Stratigraphic Unit 6, Upper Terrace, Fig. 1 and fig. S1), associated with an Early Levantine Mousterian (Tabun D type) stone-tool assemblages (*9*, *10*). Misliya-1 preserves much of the alveolar and zygomatic processes, part of the palate and nasal floor, and the complete left dentition from the first incisor (represented by a broken root only) to the third molar (Fig. 2A).

Three independent numerical dating methods—U-series (U-Th), combined uranium series and electron spin resonance (US-ESR) series, and thermoluminescence (TL)—carried out in three different dating laboratories yielded consistent results (Fig. 2B, figs. S2 and S3, and tables S1 and S3). A series of nine TL dates on burnt flints from Square L10 and N12 in the vicinity of the human fossil (Fig. 1, A and B) provided a mean age of 179 ± 48 thousand years (ky) (2σ) (range = 212 to 140 ky) (*11*). U-Th analyses of the dentine of the I^2 from the maxilla and of the crust adhering directly to the maxilla yielded a minimum age of 70.2 ± 1.6 ky (2σ; table S1) and 185 ± 8.0 ky (2σ; Fig. 2B and table S2), respectively (*9*). The combined US-ESR dating of the enamel of the same tooth yielded a maximum age of 174 ± 20 ky (2σ)

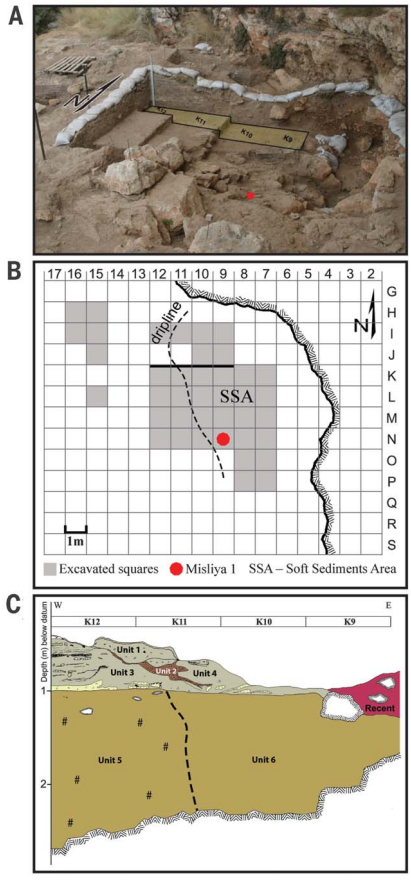


Fig. 1. The Misliya Cave excavation area at the Upper Terrace of the cave and the stratigraphy. (**A**) The excavation area and the location of the Misliya-1 maxilla (red dot). Squares K9 to K12 are indicated. (**B**) Map of the Misliya Cave Upper Terrace excavations (1 m^2 grid) with denoted excavated squares and showing the location of the human maxilla (Misliya-1). (**C**) Stratigraphic section of the Upper Terrace, squares K9 to K12. Apart from Unit 2, a Terra Rosa soil intrusion, all units contain EMP finds or assemblages. The present-day dripline roughly separates between highly cemented (Units 1,3,5) and more loosely cemented (Units 4 and 6) sediments. Misliya-1 was retrieved from the upper part of Unit 6.

[1]Department of Anatomy and Anthropology, Sackler Faculty of Medicine, Tel Aviv University, Post Office Box 39040, Tel Aviv 6997801, Israel. [2]The Dan David Center for Human Evolution and Biohistory Research and The Shmunis Family Anthropology Institute, The Steinhardt Museum of Natural History, Tel Aviv University, Post Office Box 39040, Tel Aviv 6997801, Israel. [3]Department of Anthropology and Core Facility for Micro-Computed Tomography, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria. [4]Department of Anthropology, Binghamton University (SUNY), Binghamton, NY 13902-6000, USA. [5]Centro UCM-ISCIII de Evolución y Comportamiento Humanos, Avda. Monforte de Lemos, 5, 28029, Madrid, Spain. [6]Division of Anthropology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024-5192, USA. [7]Australian Research Centre for Human Evolution (ARCHE), Environmental Futures Research Institute, Griffith University, Nathan QLD 4111, Australia. [8]National Research Center on Human Evolution (CENIEH), Paseo de la Sierra de Atapuerca 3, 09002, Burgos, Spain. [9]Research School of Earth Sciences, The Australian National University, Canberra ACT 2601, Australia. [10]Geological Survey of Israel, 30 Malkhe Israel Street, Jerusalem 9550161, Israel. [11]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, avenue de la terrasse, 91198 Gif sur Yvette Cedex, France. [12]Institut de Recherche sur les Archéomatériaux, UMR 5060 CNRS - Université de Bordeaux Montaigne, Centre de Recherche en Physique Appliquée à l'Archéologie (CRP2A), Maison de l'archéologie, 33607 PESSAC Cedex, France. [13]Departamento de Geodinámica, Estratigrafía y Paleontología, Facultad de Ciencias Geológicas, Universidad Complutense de Madrid, Ciudad Universitaria s/n, 28040, Madrid, Spain. [14]Department of Anthropology, University College London, 14 Taviton Street, London, WC1H 0BW, UK. [15]Institute of Evolutionary Medicine, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland. [16]UMR5189 PACEA Université de Bordeaux, CNRS MCC, France. [17]Department of Oral biology and Orthodontics, the Maurice and Gabriela Goldschleger School of Dental Medicine, Sackler Faculty of Medicine, Tel Aviv University, 6997801 Israel. [18]Departamento de Historia, Geografía y Comunicación, Universidad de Burgos, Facultad de Humanidades y Comunicación, 09001, Burgos, Spain. [19]Facultade de Humanidades, Universidad Isabel I, Spain. [20]Área de Prehistoria, Universitat Rovira i Virgili, Avinguda Catalunya 35, 43002 Tarragona, Spain. [21]Institut Català de Paleoecologia Humana i Evolució Social (IPHES), Marcel·lí Domingo s/n, 43007 Tarragona, Spain. [22]Institute of Earth Science, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel [23]Department of Maritime Civilizations, Recanati Institute of Maritime Studies, University of Haifa, Haifa, Mount Carmel 3498838, Israel. [24]Sonia and Marco Nadler Institute of Archaeology Tel Aviv University, Tel Aviv 69978, Israel. [25]Peabody Museum of Archaeology and Ethnology, Harvard University, 11 Divinity Avenue, Cambridge, MA 02138, USA. [26]Department of Paleoanthropology, Institute of Paleontology and Paleoanthropology, Chinese Academy of Science, Str. Xizhimenwai no. 144, 100044 Beijing, China. [27]Department of Radiology, Carmel Medical Center, Haifa, 3436212 Israel. [28]Zinman Institute of Archaeology, University of Haifa, Haifa, Mount Carmel 3498838, Israel. [29]Institute of Archaeology, The Hebrew University of Jerusalem, Jerusalem 9190501, Israel.
***Corresponding author. Email: anatom2@tauex.tau.ac.il** †These authors contributed equally to this work.

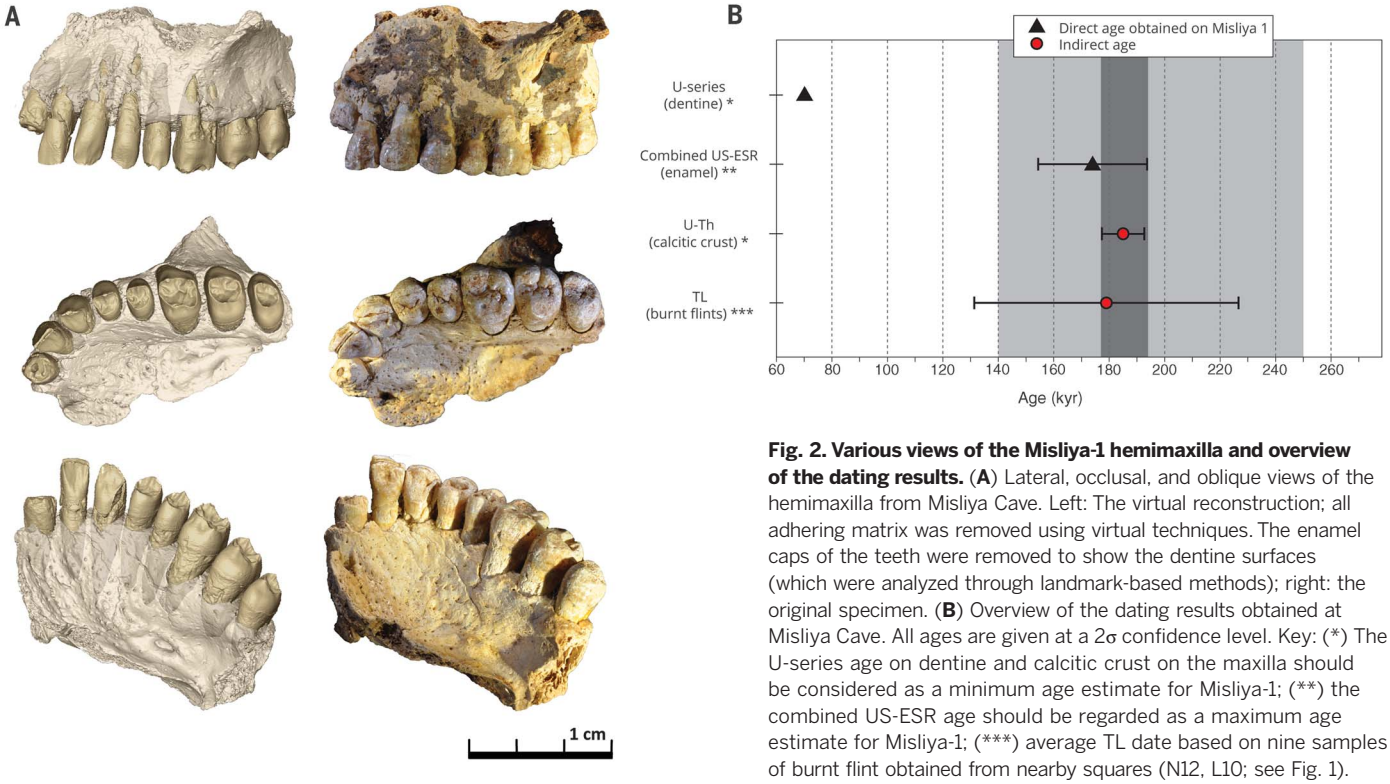


Fig. 2. Various views of the Misliya-1 hemimaxilla and overview of the dating results. (**A**) Lateral, occlusal, and oblique views of the hemimaxilla from Misliya Cave. Left: The virtual reconstruction; all adhering matrix was removed using virtual techniques. The enamel caps of the teeth were removed to show the dentine surfaces (which were analyzed through landmark-based methods); right: the original specimen. (**B**) Overview of the dating results obtained at Misliya Cave. All ages are given at a 2σ confidence level. Key: (*) The U-series age on dentine and calcitic crust on the maxilla should be considered as a minimum age estimate for Misliya-1; (**) the combined US-ESR age should be regarded as a maximum age estimate for Misliya-1; (***) average TL date based on nine samples of burnt flint obtained from nearby squares (N12, L10; see Fig. 1). Dark gray: Age range for Misliya-1, based on dates obtained from the fossil (U-Th provides the minimum age and combined US-ESR the maximum age), is between 177 ky (=185 – 8 ky) to 194 ky (=174 + 20 ky). Light gray: Age range for the EMP period in the Levant (250 to 140 ky) based on the combination of TL dates obtained for Tabun Cave (*13*), Hayonim Cave (*12*), and Misliya Cave (*11*).

(Fig. 2B, fig. S3, and tables S1 and S3) (*9*). All these dates, except for the U-series dating of the dentine, which exclude the possibility of recent intrusion, fall within the time range for the Early Levantine Mousterian lithic industry (Tabun D-type) observed at Tabun, Hayonim, and Misliya caves (i.e., ~250 to ~140 ky) (*11*–*13*) and are older than the upper range defined for the EMP sequence in Misliya Cave (>165 ky) (*11*). Collectively, the evidence suggests an early marine isotope stage 6 (MIS 6) age for the Misliya-1 fossil. The age range for Misliya-1, based on dates directly connected with the fossil (U-Th on crust providing the minimum boundary and the maximum boundary of US-ESR on the enamel of I^2), is between 177 and 194 ky [for details and calculation methods, see (*9*) and fig. S2].

The insertion of the zygomatic root in Misliya-1 is relatively anteriorly placed, at the level of M^1, similar to recent *H. sapiens* as well as the fossils from Herto and Jebel Irhoud (*1*, *3*). The zygomaticoalveolar crest is strongly curved and inserts at a low position relative to the dentition. Three-dimensional (3D) geometric morphometric (GM) analysis (*9*) of the maxillary morphology (Fig. 3A) shows that Misliya-1 is similar to modern humans and is most dissimilar to Neandertals and some Middle Pleistocene hominins. Based on the logarithm of the centroid size of the 3D data (*9*), Misliya-1 is smaller than all fossil *Homo* specimens in our sample and falls in the range of variation observed for recent modern humans. In addition, the anterior placement of the incisive foramen, the sloped or level nasal floor configuration, and the shape of the dental arcade in Misliya-1 (Fig. 2A) are characteristic of modern humans, although individual features can occasionally be found in earlier taxa.

The I^2 (Fig. 2A and fig. S4) shows several features that are characteristic of *H. sapiens* (*14*), including a flat labial surface [labial convexity grades 0 or 1, scoring based on Arizona State University Dental Anthropology System (ASUDAS)], a straight incisal edge, very slight shoveling (ASUDAS grade 1), and lack of a lingual tubercle. The presence of a lingual groove can also be found in Pleistocene *H. sapiens* samples such as Qafzeh and Huanglong. The canine lacks the mass-additive traits typical of Asian *Homo erectus* (*15*), Middle Pleistocene European specimens, and Neandertals, and resembles Qafzeh and Skhul. Unlike Neandertals canines, the shovel shape is not pronounced, and there is no lingual tubercle or mesial canine ridge (Fig. 2A).

The upper premolars display relatively simple occlusal surfaces and lack accessory marginal tubercles and buccal cingulum (Fig. 2A). The Misliya-1 premolars display the typical high and narrow crown of *H. sapiens*. In the occlusal view, the P^3 shows a slight lingual narrowing, which is less pronounced in the P^4. This contrasts with the characteristic Neandertal pattern featuring a low and broad crown and subequal buccal and lingual aspects of the crown in both upper premolars. The proportion of occlusal area (defined by the occlusal rim) is large relative to the crown base area in the upper premolars of Misliya-1, unlike in Neandertals, where the occlusal area appears compressed relative to the crown base area. This compression in Neandertal upper premolars is homologous to the relative reduction of the occlusal polygon found in Neandertal M^1s (*16*, *17*), and this latter feature is absent in Misliya-1.

The Misliya-1 maxillary teeth are within the upper size range of modern humans (table S5). Size proportions between the anterior and posterior teeth differentiate Misliya-1 from Neandertals (fig. S5). The buccolingual (BL) size ratio of the I^2 to M^1 in Misliya-1 (62.6) is just outside the upper limit of the range of modern humans (mean = 55.6, SD = 3.4, n = 31, range = 48.2 to 62.5), is similar to the mean of Qafzeh and Skhul (mean = 63.4, SD = 4.9, n = 9, range = 56.1 to 71.4), and well below the lower limit of the range of Neandertals (mean = 70.3, SD = 3.1, n = 13, range = 66.7 to 76.0). Therefore, Misliya-1 does not exhibit the relative expansion of the anterior dentition characteristic of Neandertals (*18*). Tooth root size and morphology are also within the range of modern humans (fig. S4).

Two-dimensional GM analysis (*9*) of the M^1 crown outline (Fig. 4) reveals that Misliya-1 is separate from Neandertals and other European Middle Pleistocene hominins, placing it with modern humans and near to Jebel Irhoud. It differs from Neandertals as well as from other European Middle Pleistocene fossils by not displaying the skewed rhomboidal crown outline

and large and protruding hypocone. The relative sizes of the M^1 protocone and hypocone align Misliya-1 with modern humans and differentiate it from Neandertals (table S4).

The 3D GM analysis (*9*) of the premolars, including the enamel-dentine junction (EDJ) occlusal area and cementum-enamel junction (CEJ) (fig. S6), shows that the Misliya-1 premolars are located in quadrants exclusively occupied by *H. sapiens*, with the exception of one Atapuerca Sima de los Huesos (SH) P^3 (that is located in the same quadrant but far from Misliya-1) and the P^4 of Amud 1. A similar analysis of the Misliya-1 M^2 (Fig. 3B) places it in an area exclusively occupied by contemporary *H. sapiens* (and the Liujiang specimen), which are characterized by a reduction of the hypocone and a buccolingually widened (rectangular) crown base. This contrasts with what is observed in Neandertals and most other European Middle Pleistocene fossils where the hypocone is relatively more developed. The Qafzeh specimens are quite variable but uniformly display a larger hypocone than does Misliya-1. The strong reduction of the hypocone observed in M^1 (table S4), M^2, and M^3 of Misliya-1 is most frequently observed in *H. sapiens*, although it can occasionally be found in other *Homo* groups (*17*, *18*).

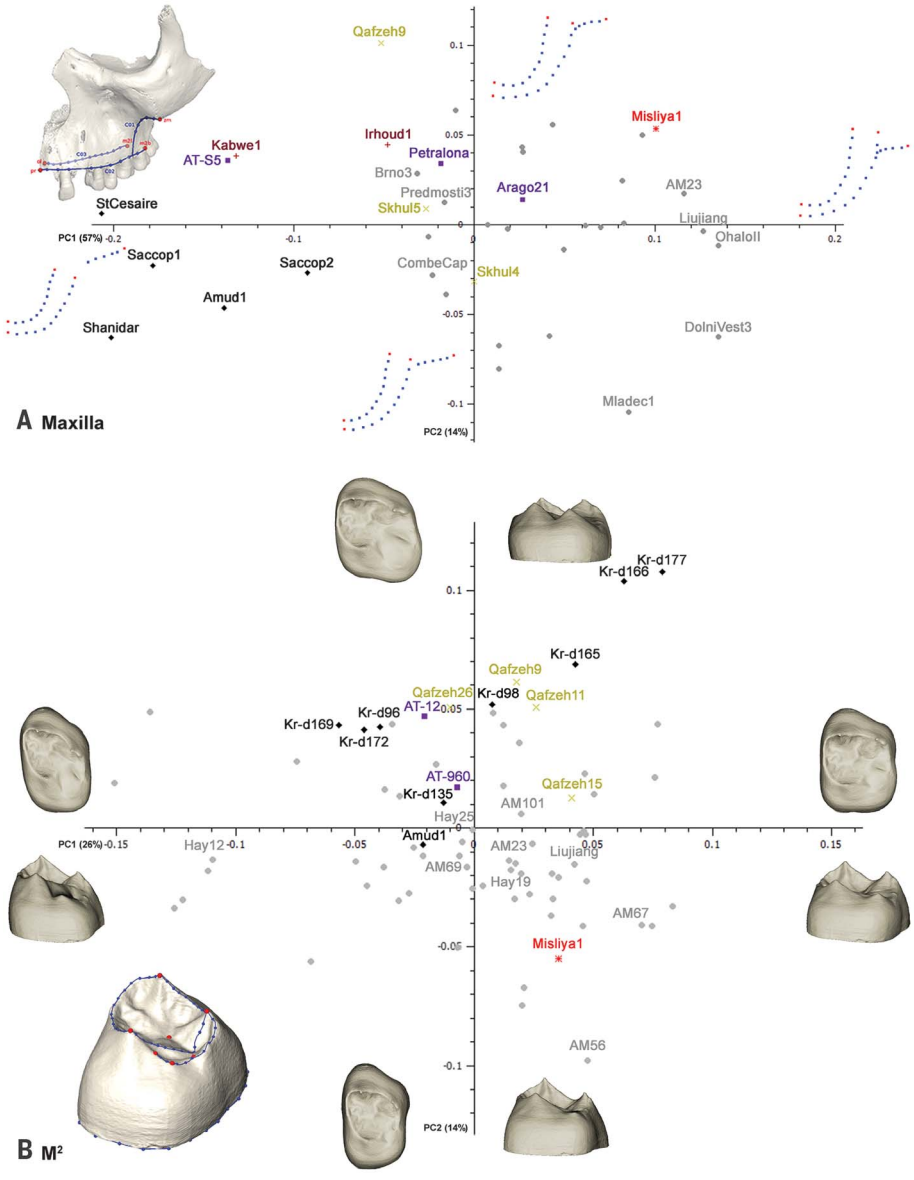


Fig. 3. First two principal components (PCs) in shape space and associated warped surfaces at the extremities of the axes. Noteworthy shape changes in the warpings are as follows: (**A**) Maxillary bone. PC1, anterior-posterior position and angle of the zygomatic root. PC2, curvature of the zygomatico-alveolar crest. (**B**) Upper M^2; PC2, relative expansion and height of the hypocone; buccolingual relative size of the EDJ occlusal area to the crown base. On the left side of each of the plots, the landmark configurations used for the respective anatomical unit are represented (landmarks in red, curve semilandmarks and pseudolandmarks on CEJ in blue). Prosthion (pr), orale (ol), zygomaxillare (zm), midpoint of the M^2 alveolar socket buccally (m2b) and lingually (m2l); upper region of the zygomaticoalveolar crest (C01); buccal alveolar margin (C02), lingual alveolar margin (C03). Red star, Misliya-1; gray circles, recent modern humans (without labels), Upper Paleolithic and Epipaleolithic modern *Homo* (with labels); black diamonds, Neandertals; yellow X, early modern humans; violet square, European Early and Middle Pleistocene *Homo*; burgundy plus sign, African Early and Middle Pleistocene *Homo*; for the specimen labels, refer to table S7.

Overall, the Misliya-1 teeth are distinct from those of the Middle Pleistocene specimens from Europe, Africa, and Asia such as Atapuerca (SH), Steinheim, Rabat, Qesem Cave, Chaoxian, and Xujiayao. Although some dental features seen in Misliya-1 can occasionally also be found in some of these samples, the entire suite of metric and morphological traits seen in the Misliya-1 maxillary bone and teeth is more consistent with *H. sapiens* than with Neandertals or other Middle Pleistocene hominin groups. Indeed, the combination of features in the incisor and canine appears to occur only in *H. sapiens* (*19*).

Middle Pleistocene fossils from southwest Asia (e.g., Qesem Cave, Zuttiyeh) are rare and display a mixture of features considered characteristic of Neandertals or modern humans, thus complicating their taxonomic assignment (*20*–*22*). Although incomplete, the Misliya-1 maxilla does not exhibit any derived skeletal or dental Neandertal features. A specific comparison with the earlier teeth from Qesem Cave (*20*, *21*) reveals a number of differences. Specifically, the Qesem I^2 shows a pronounced lingual tubercle, greater degree of labial curvature, and more pronounced shoveling, whereas the Qesem C^1 shows more pronounced shoveling, a lingual tubercle, and a canine mesial ridge. All of these features are more commonly found in Neandertal anterior teeth and represent points of departure from the morphology seen in Misliya-1 teeth. In contrast, Misliya-1 resembles the later Levantine *H. sapiens* fossils from the sites of Skhul and Qafzeh regarding many dental features, but it also differs from them regarding the degree of hypocone reduction seen in Misliya-1.

The geographical origin, timing, and identification of the last common ancestor of Neandertals and modern humans remain controversial (*23*, *24*). Nevertheless, the evolutionary emergence of Neandertals in Europe from their Middle Pleistocene precursors [e.g., Atapuerca (SH), Steinheim, Ehringsdorf] is better established, despite the possibility that more than one lineage coexisted in the European Middle Pleistocene (*25*). The geographical origin of *H. sapiens* is generally considered to be Africa, and the Jebel Irhoud fossils, recently dated to ~300 ky ago (*2*), are thought to represent an "early phase of *H. sapiens* evolution" [(*1*), p. 291]. Younger fossils from the sites of Omo (~195 ky ago) and Herto (~160 ky ago) have been attributed to *H. sapiens* (*3*, *4*). Nevertheless, the African fossil records reveal temporal overlaps between more "archaic" and more "modern" forms of early *H. sapiens* (*24*). These African specimens are thought to be members of the *H. sapiens* clade, even though some of them fall outside the range

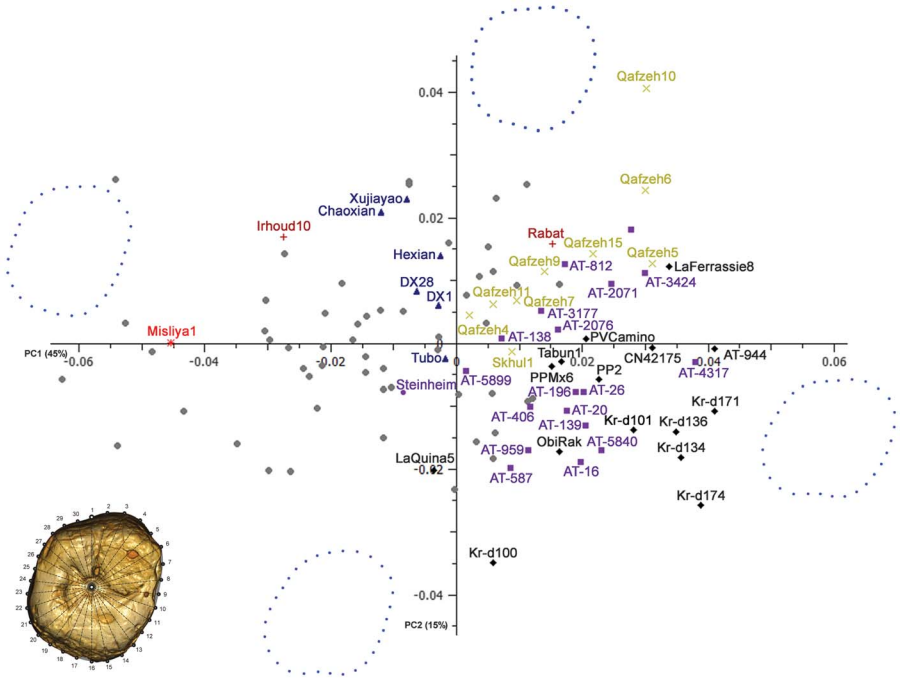


Fig. 4. First two principal components (PCs) of the crown shape of Misliya-1 M[1]. Misliya-1 is distinct from Neandertals and other Middle Pleistocene hominins and clearly grouped with modern humans. Red star, Misliya-1; gray circles, recent modern humans (without labels), Upper Paleolithic and Epipaleolithic modern *Homo* (with labels); black diamonds, Neandertals; yellow X, early modern humans; violet square, European Early and Middle Pleistocene *Homo*; burgundy plus sign, African Early and Middle Pleistocene *Homo*; blue triangle, Middle Pleistocene Asian specimens; for the specimen labels, refer to table S7.

of variation of Holocene humans regarding certain features (*3*, *24*). Similarly, many of the teeth, which are thought to represent early *H. sapiens* from North Africa, retain primitive features (*26*).

Misliya-1 considerably pushes back the timing of the earliest migration of members of the *H. sapiens* clade out of Africa, well predating Qafzeh and Skhul in the Levant, and Daoxian and Liujiang in China [(*8*) and (*27*); but see (*28*)].

Archaeologically, the EMP layers of Misliya cave document the emergence of novel technological concepts in the Levant, including full-fledged Levallois technology and laminar technology (*29*). Similar technological concepts have been documented at contemporary and earlier Middle Stone Age sites in Africa, i.e., the Maghreb (Jebel Irhoud), eastern Africa (Gademotta and Kulkuletti formations, Ethiopia, and the Kapthurin Formation, Kenya), and southern Africa (Kathu Pan) (*2*, *30*–*34*). Thus, similar to the recent findings from Jebel Irhoud (*1*, *2*), the evidence from Misliya Cave suggests that the emergence of full-fledged Levallois technology in the Levant may have also been associated with the occurrence of *H. sapiens*.

The region of southwest Asia represents a major biogeographic corridor for hominin migrations during our evolutionary history. Given the geographical proximity of the Levant to Africa, it is possible that the dispersals documented at Misliya Cave (177 to 194 ky ago), Qafzeh and Skhul Caves (90 to 120 ky ago), and Manot Cave (50 to 60 ky ago) reflect expansions of the geographical range of *H. sapiens*, fluctuating in response to demographic or environmental factors (*35*).

To date, Misliya-1 appears to represent the earliest fossil evidence of the migration of members of the *H. sapiens* clade out of Africa. It therefore opens the door to the possibility that *H. sapiens* dispersal from Africa could have occurred earlier than previously thought (probably before 200 ky ago), as has been recently suggested based on genetic evidence (*36*).

REFERENCES AND NOTES

1. J.-J. Hublin *et al.*, *Nature* **546**, 289–292 (2017).
2. D. Richter *et al.*, *Nature* **546**, 293–296 (2017).
3. T. D. White *et al.*, *Nature* **423**, 742–747 (2003).
4. I. McDougall, F. H. Brown, J. G. Fleagle, *Nature* **433**, 733–736 (2005).
5. N. Mercier *et al.*, *J. Archaeol. Sci.* **20**, 169–174 (1993).
6. C. B. Stringer, R. Grün, H. P. Schwarcz, P. Goldberg, *Nature* **338**, 756–758 (1989).
7. H. P. Schwarcz *et al.*, *J. Hum. Evol.* **17**, 733–737 (1988).
8. W. Liu *et al.*, *Nature* **526**, 696–699 (2015).
9. Materials and methods are available as supplementary materials online.
10. Y. Zaidner, M. Weinstein-Evron, *Before Farm.* **2012**, 1–23 (2012).
11. H. Valladas *et al.*, *J. Hum. Evol.* **65**, 585–593 (2013).
12. N. Mercier *et al.*, *J. Archaeol. Sci.* **34**, 1064–1077 (2007).
13. N. Mercier, H. Valladas, *J. Hum. Evol.* **45**, 401–409 (2003).
14. M. Martinón-Torres *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13279–13282 (2007).
15. F. Weidenreich, *The dentition of* Sinanthropus pekinensis; *a comparative odontography of the hominids. Palaeontol Sinica, New Series D1* (Peking, 1937).
16. S. E. Bailey, *J. Hum. Evol.* **47**, 183–198 (2004).
17. M. Martinón-Torres, J. M. Bermúdez de Castro, A. Gómez-Robles, L. Prado-Simón, J. L. Arsuaga, *J. Hum. Evol.* **62**, 7–58 (2012).
18. J. M. Bermúdez de Castro, M. E. Nicolas, *Am. J. Phys. Anthropol.* **96**, 335–356 (1995).
19. S. E. Bailey, J. Hublin, in *Anthropological Perspectives on Tooth Morphology: Genetics, Evolution, Variation*, G. Scott, J. Irish, Eds. (Cambridge Univ. Press, Cambridge, 2013), pp. 222–249.
20. G. Weber *et al.*, *Quat. Int.* **398**, 159–174 (2016).
21. I. Hershkovitz *et al.*, *Am. J. Phys. Anthropol.* **144**, 575–592 (2011).
22. S. E. Freidline, P. Gunz, I. Janković, K. Harvati, J.-J. Hublin, *J. Hum. Evol.* **62**, 225–241 (2012).
23. A. Gómez-Robles, J. M. Bermúdez de Castro, J.-L. Arsuaga, E. Carbonell, P. D. Polly, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 18196–18201 (2013).
24. C. Stringer, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **371**, 1698 (2016).
25. J. L. Arsuaga *et al.*, *Science* **344**, 1358–1363 (2014).
26. S. E. Bailey, T. D. Weaver, J. Hublin, in *Human Paleontology and Prehistory*, A. Marom, E. Hovers, Eds. (Springer International, Cham, Switzerland, 2017), pp. 215–232.
27. G. Shen *et al.*, *J. Hum. Evol.* **43**, 817–829 (2002).
28. W. Liu, X. Wu, L. Wang, *Acta Anthropol. Sin.* **25**, 177–194 (2006).
29. M. Weinstein-Evron, Y. Zaidner, A. Tsatskin, R. Yeshurun, I. Hershkovitz, in *Quaternary of the Levant: Environments, Climate Change, and Humans*, Y. Enzel, O. Bar-Yosef, Eds. (Cambridge Univ. Press, Cambridge, 2017), pp. 225–230.
30. K. Douze, A. Delagnes, *J. Hum. Evol.* **91**, 93–121 (2016).
31. C. R. Johnson, S. McBrearty, *J. Hum. Evol.* **58**, 193–200 (2010).
32. N. Porat *et al.*, *J. Archaeol. Sci.* **37**, 269–283 (2010).
33. Y. Sahle, L. E. Morgan, D. R. Braun, B. Atnafu, W. K. Hutchings, *Quat. Int.* **331**, 6–19 (2014).
34. C. A. Tryon, S. McBrearty, P.-J. Texier, *Afr. Archaeol. Rev.* **22**, 199–229 (2005).
35. I. Hershkovitz *et al.*, *Nature* **520**, 216–219 (2015).
36. C. Posth *et al.*, *Nat. Commun.* **8**, 16046 (2017).

ACKNOWLEDGMENTS

Field work in Misliya Cave was supported by the Dan David Foundation, the Irene Levi-Sala CARE Archaeological Foundation, the Leakey Foundation, the Thyssen Foundation, and the Faculty of Humanities of the University of Haifa. Laboratory work and dating were supported by the Israel Science Foundation (grant no. 1104/12). The anthropological study was supported by the Dan David Foundation, Ministerio de Economía y Competitividad of Spain (CGL2015-65387-C3-2-3-P MINECO/FEDER), Fundación Atapuerca, and The Leakey Foundation. The ESR dating study received funding from the Marie Curie International Outgoing Fellowship (IOF) 626474 and the Australian Research Council Future Fellowship FT150100215. N.M. is grateful to LaScArBx ANR-10-LABX-52 for support. Work on the virtual specimens was supported by the Life Science Faculty University of Vienna; Oesterreichische Nationalbank, Anniversary Fund, project no. 16121; the Swiss National Science Foundation grant nos. 31003A 156299/1 and 31003A 176319; A.E.R.S. Dental Medicine Organizations GmbH, Vienna, Austria, project no. FA547014; and the Siegfried Ludwig–Rudolf Slavicek Foundation, Vienna, Austria, project no. FA547016. Special thanks are due to the late Dan David and his son, Ariel David, for their inspiration and financial support throughout the years. Thanks also go to O. Bar-Yosef (Harvard University) and A. M. Tillier (University of Bordeaux), who read and commented on a previous version of this paper. All data generated in this study are in the supplementary materials. Permission to study the fossil (Misliya-1) can be obtained from the corresponding author. Micro-CT images are available from G.W.W. under a material transfer agreement with I.H. of Tel Aviv University (Misliya-1). Full acknowledgments are in the supplementary materials. I.H. and M.W.-E. direct the Misliya Cave research project; I.H., G.W.W., R.Q., J.L.A., M.M.-T., J.M.B.dC., L.M.-F., P.S., R.S., H.M., V.S., L.R., R.G., C.L., J.M.C., Y.C., X.W., and N.P. carried out various aspects of the anthropological study of the Misliya human remains; G.W.W., C.F., V.A.K., and R.S carried out the virtual image manipulation and geometric morphometric investigations; Y.Z. and M.W.-E. conducted the archaeological studies at the cave; A.A and M.B.-M. conducted the U/Th analysis; H.V. and N.M. conducted the TL dating of the archaeological layers; M.D., R.G., and L.K. conducted the U-series and ESR dating of the fossil remains; A.F., R.S.-G., and A.T. conducted the geoarchaeological and geomorphological studies of the cave; L.W. and R.Y. conducted the study of the faunal remains; D.B.-Y.M. studied the mollusc shells; and I.G.-Y. carried out the use-wear analysis. All authors participated in compiling the manuscript. The authors declare no competing financial interests.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6374/456/suppl/DC1
Materials and Methods
Supplementary Text
Acknowledgments
Figs. S1 to S6
Tables S1 to S7
References (*37*–*114*)

31 August 2017; accepted 21 December 2017
10.1126/science.aap8369

CORAL REEFS

Plastic waste associated with disease on coral reefs

Joleah B. Lamb,[1,2,3]* Bette L. Willis,[2,3] Evan A. Fiorenza,[1,4] Courtney S. Couch,[1,5,6] Robert Howard,[7] Douglas N. Rader,[8] James D. True,[9] Lisa A. Kelly,[3,10] Awaludinnoer Ahmad,[11,12] Jamaluddin Jompa,[12] C. Drew Harvell[1]

Plastic waste can promote microbial colonization by pathogens implicated in outbreaks of disease in the ocean. We assessed the influence of plastic waste on disease risk in 124,000 reef-building corals from 159 reefs in the Asia-Pacific region. The likelihood of disease increases from 4% to 89% when corals are in contact with plastic. Structurally complex corals are eight times more likely to be affected by plastic, suggesting that microhabitats for reef-associated organisms and valuable fisheries will be disproportionately affected. Plastic levels on coral reefs correspond to estimates of terrestrial mismanaged plastic waste entering the ocean. We estimate that 11.1 billion plastic items are entangled on coral reefs across the Asia-Pacific and project this number to increase 40% by 2025. Plastic waste management is critical for reducing diseases that threaten ecosystem health and human livelihoods.

Outbreaks of disease on coral reefs threaten one of the most biodiverse ecosystems on the planet (*1*), jeopardizing the U.S. $375 billion in goods and services that they provide to people each year through fisheries, tourism, and coastal protection (*2*). Plastic waste can host pathogens that are frequently implicated as triggers of disease outbreaks on coral reefs (*3*–*9*). For example, microbial communities colonizing polypropylene marine debris were dominated by the genus *Vibrio* (*10*), an opportunistic pathogenic bacteria of a globally devastating group of coral diseases known as white syndromes (*11*). Although an estimated 4.8 million to 12.7 million metric tons of plastic waste enter the ocean in a single year (*12*), the resulting influence on disease susceptibility in the marine environment is unknown. Microbial rafting on plastic debris has been shown to strongly control surface longevity (*13*) and is highest in tropical regions near the equator compared with more polar regions (*14*), suggesting that coral reef ecosystems could have high levels of colonized plastic waste.

We surveyed 159 coral reefs spanning eight latitudinal regions from four countries in the Asia-Pacific for plastic waste and evaluated the influence of plastic on diseases that affect keystone reef-building corals (*15*) (benthic area = 12,840 m^2) (Fig. 1). The Asia-Pacific region contains 55.5% of global coral reefs (*2*) and encompasses 73.0% of the global human population residing within 50 km of a coast (*12*) (table S1). Overall, we documented benthic plastic waste (defined as an item with a diameter >50 mm) on one-third of the coral reefs surveyed, amounting to 2.0 to 10.9 plastic items per 100 m^2 of reef area [95% confidence interval (CI), n = 8 survey regions]. The number of plastic items observed on each reef varied markedly among countries, from

[1]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA. [2]Australian Research Council Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, Queensland 4811, Australia. [3]College of Science and Engineering, James Cook University, Townsville, Queensland, Australia. [4]School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA 98195, USA. [5]Hawaii Institute of Marine Biology (HIMB), University of Hawaii at Manoa, Kaneohe, HI 96744, USA. [6]Ecosystem Sciences Division, National Oceanic and Atmospheric Administration (NOAA) Pacific Islands Fisheries Science Center, Honolulu, HI 96818, USA. [7]Marine Programme, Fauna & Flora International, Yangon, Myanmar. [8]Oceans Program, Environmental Defense Fund, New York, NY 10010, USA. [9]Center for Biodiversity in Peninsular Thailand, Prince of Songkla University, Hat Yai, Songkhla, Thailand. [10]Fish Ecology and Conservation Physiology Laboratory, Department of Biology and Institute of Environmental Science, Carleton University, Ottawa, Ontario, Canada. [11]The Nature Conservancy, Raja Ampat Field Office, North Sorong, West Papua, Indonesia. [12]Faculty of Marine Science and Fisheries, Hasanuddin University, Makassar, South Sulawesi, Indonesia.
*Corresponding author. Email: joleah.lamb@cornell.edu

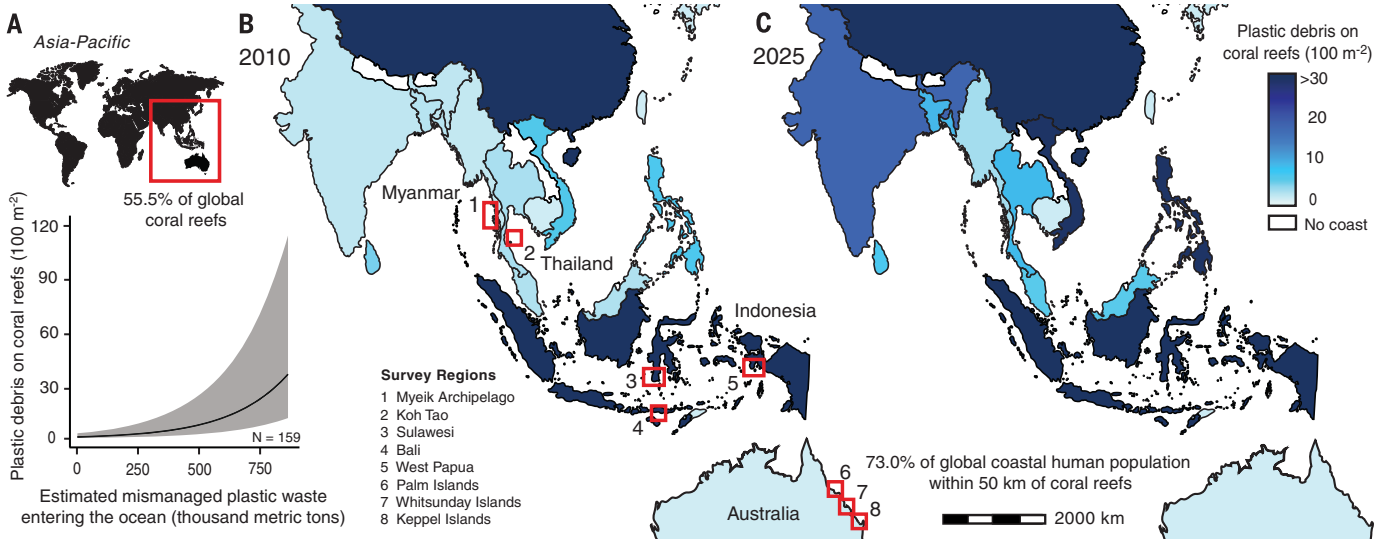


Fig. 1. Estimated plastic debris levels on coral reefs. (**A**) Modeled association between plastic debris on coral reefs from surveys of 159 reefs in eight regions [red squares in (B)] from 2011–2014 and estimated levels of mismanaged plastic waste (thousand metric tons), assuming that 25% of waste entered the ocean in 2010 from human populations living within 50 km of the coast for each country (*12*). Reef locations can be found in table S14. The Asia-Pacific region encompasses 9 of 10 countries with the highest global levels of estimated mismanaged plastic waste entering the ocean (table S3). Gray shading represents the upper and lower 95% CIs for the model. (**B** and **C**) Modeled plastic debris levels on coral reefs (100 m^{-2}), as projected using the association between estimated mismanaged plastic waste entering the ocean in 2010 for each sovereign country (*12*) and plastic debris surveys between 2011–2014 [shown in (A)]. The color scale represents the minima and maxima model estimates of mismanaged plastic waste on coral reefs from 2010 (table S5). Projections of plastic debris on coral reefs for Indonesia and China in 2025 were set to the maxima from 2010, owing to the limitations of the model range. Countries without a coastline are shown in white.

maxima in Indonesia [25.6 items per 100 ± 12.2 m^2 (here and elsewhere, the number after the ± symbol denotes SEM)] to minima in Australia (0.4 items per 100 ± 0.3 m^2) (table S2).

Terrestrially derived pollutants have been implicated in several disease outbreaks in the ocean (*16*). However, no studies have examined the influence of plastic waste on disease risk in a marine organism. In this work, we visually examined 124,884 reef-building corals for signs of tissue loss characteristic of active disease lesions (*15*) (fig. S1). We found plastic debris on 17 genera from eight families of reef-forming corals. When corals were not in contact with plastic debris, the likelihood of disease was 4.4 ± 0.2% across all eight regions (range = 2.8 to 8.4%, generalized linear mixed model, likelihood ratio test among regions: $\chi^2_7 = 10.382$, $P = 0.168$) (Fig. 2A). In contrast, in the presence of plastic debris, the likelihood of disease occurrence in corals increased significantly by more than a factor of 20 to 89.1 ± 3.2% (generalized linear mixed model: z score = 27.24, $P < 0.001$, $n = 331$ transects) (Fig. 2B and table S3).

Human population size in coastal regions and the quality of waste management systems largely determine which countries contribute the greatest plastic loads entering the ocean, given that an estimated 80% of marine plastic debris originates from land (*12*). Accordingly, we modeled the relationship between our documented levels of plastic debris on coral reefs (n = 437 transects from Australia, Myanmar, Thailand, and Indonesia) and Jambeck *et al.*'s estimated levels of mismanaged plastic waste entering the ocean (*12*) from these four countries in 2010 (*15*) (generalized linear mixed model: Akaike information criterion = 662.3, $z = 3.95$, $P < 0.001$) (Fig. 1A, fig. S2, and table S4). Our model encompasses the range of mismanaged plastic waste entering the ocean introduced by coastal populations from 15 of the 17 (88%) sovereign countries in the Asia-Pacific region (maximum = 804,214 metric tons, minimum = 3472 metric tons), of which 9 are among the top 10 plastic-polluting countries globally (table S5). Assuming that improvements in waste management infrastructure did not occur during our survey period (2011–2014) and that the plastic waste emanated from adjacent terrestrial point sources, we estimate that levels of plastic debris on coral reefs for each country in the Asia-Pacific ranged from 0.9 to 26.6 plastic items per 100 m^2 in 2010 (95% CI) (Fig. 1B and table S5). This amounts to an estimated 11.1 billion items of plastic on coral reefs across the Asia-Pacific (95% CI = 1.2 billion to 105.5 billion items, n = 15 countries), which is likely underestimated owing to the exclusion of China and Singapore because they fall outside of the model range (table S5).

By 2025, the cumulative quantity of plastic waste potentially entering the marine environment from land is predicted to increase by one order of magnitude (*12*). Using this projection and assuming that the area encompassed by coral reefs remains constant, we estimate that 15.7 billion plastic items will be entangled on coral reefs across the Asia-Pacific by 2025 (the "business-as-usual" scenario for global infrastructure: 95% CI = 1.7 billion to 149.2 billion items) (Fig. 1C and table S5). According to our model, the predicted geographic distribution of plastic debris on coral reefs does not change substantially between now and 2025 (Fig. 1, B and C), but the disparity in quantities of accumulated plastic waste between developing and industrialized countries grows considerably. For example, plastic debris on coral reefs increases by only ~1% in high-income countries such as Australia but nearly doubles in a similarly populated low-income country such as Myanmar (Fig. 1C and table S5).

Comparative analyses of disease prevalence among different diseases in the presence versus absence of plastics can offer insights into potential mechanisms that increase disease susceptibility in corals. Reef-building corals in contact with plastic debris were affected by four of six common diseases globally (*17*), whereas corals without plastic debris were affected by all six diseases but at much lower prevalence levels (table S6). Disease assemblages on reef-building corals differ distinctly when contact with plastic waste is present versus absent, as visualized by a principal coordinates ordination analysis (*15*) (permutational multivariate analysis of variance: $F = 11.86$, $P < 0.001$, $n = 75$ paired transects) (Fig. 3). In particular, three key diseases associated with rapid coral mortality increased markedly when plastic debris was in contact with coral tissues: Skeletal eroding band disease increased from 1.2 ± 0.1% to 43.9 ± 5.1% (increased likelihood = 24%), white syndromes increased from 1.9 ± 0.2% to 19.0 ± 4.0% (increased likelihood = 17%), and black band disease increased from 0.6 ± 0.1% to 14.7 ± 3.9% (increased likelihood = 5%) (tables S7 to S10).

Given the widespread distribution of plastic debris on coral reefs and the consequent increased

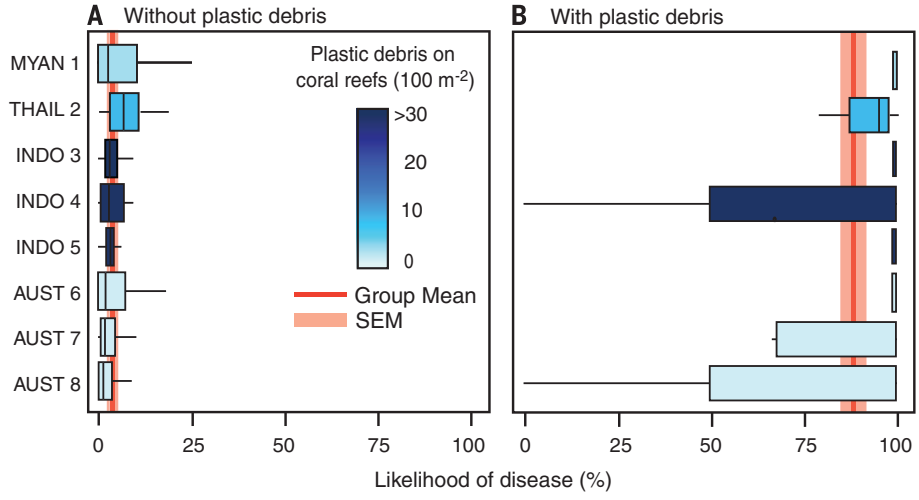


Fig. 2. Plastic waste influences disease susceptibility of reef-building corals. (**A** and **B**) Box (median and 50% quantile) and whisker (95% quantile) plots of coral disease likelihood for each of eight regions in four countries in the Asia-Pacific when no plastic waste is present (A) (n = 362 transects) and when plastic waste is present (B) (n = 75 transects). The red line represents the mean and the light red bar denotes ±1 SEM across all eight regions. Boxes are shaded according to model estimates of plastic debris on coral reefs per 100 m^2 from Fig. 1C. MYAN, Myanmar; THAIL, Thailand; INDO, Indonesia; AUST, Australia.

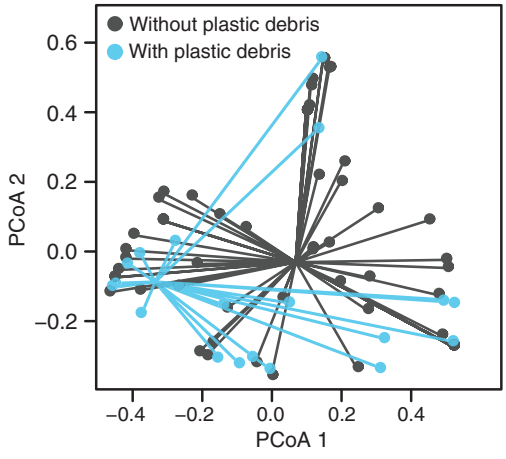


Fig. 3. Reef-building corals with plastic debris have different disease assemblages than corals without plastic debris. Multivariate spatial representation of the relative abundance and composition of coral disease assemblages, as determined by a principal coordinates analysis (PCoA) (n = 75 paired transects). Vectors for each group illustrate the median spatial distance within the group. Disease assemblages represent six diseases—skeletal eroding band, white syndromes, black band, growth anomalies, brown band, and atramentous necrosis—recorded commonly across the globe.

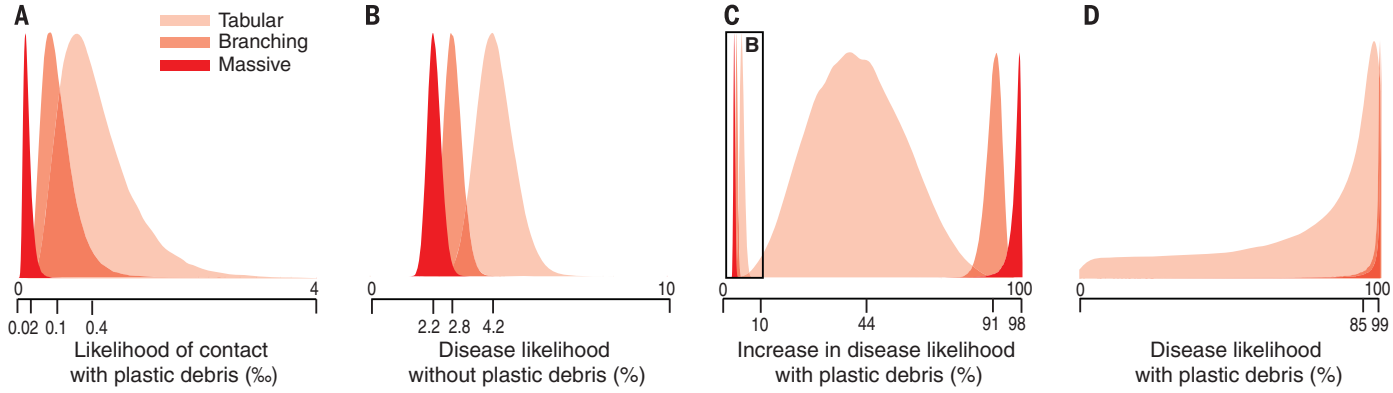


Fig. 4. Coral morphological complexity influences risk to plastic debris and disease. (**A** to **D**) Posterior probability density functions of coral species grouped into three broad morphological classifications. Structural complexity is determined by coral species (*18*); see table S11 for classifications. Minimum, maximum, and peak values are shown for each structural complexity classification group: massive (dark red), branching (medium red), and tabular (light red). For ease of comparison, the inset in (C) represents the likelihood of disease without plastic debris [as shown in (B)].

likelihood of coral mortality from disease, we evaluated the potential for plastic debris and disease to affect structural complexity provided by habitat-forming corals. The structural complexity formed by corals underpins the availability of microhabitats for coral reef–associated organisms (*18*). We grouped coral species into three broad classifications based on the increasing structural complexity of their colony morphologies (massive < branching < tabular) (table S11) and determined that plastic debris is eight times more likely to affect reef corals with greater structural complexity (tabular and branching versus massive, n = 348 transects; posterior probability functions) (Fig. 4A and table S12). Massive coral morphologies are less likely to maintain contact with plastic debris; however, they exhibit the greatest increase in disease risk when this occurs (likelihood is increased by 98%) (Fig. 4, B to D, and table S13).

Our study shows that plastic debris increases the susceptibility of reef-building corals to disease. Plastics are a previously unreported correlate of disease in the marine environment. Although the mechanisms remain to be investigated, the influence of plastic debris on disease development may differ among the three main global diseases that we observed. For example, plastic debris can cause physical injury and abrasion to coral tissues by facilitating invasion of pathogens (*19*) or by exhausting resources for immune system function during wound-healing processes (*20*). Experimental studies show that artificially inflicted wounds to corals are followed by the establishment of the ciliated protozoan *Halofolliculina corallasia*, the causative agent of skeletal eroding band disease (*21*). Plastic debris could also directly introduce resident and foreign pathogens or may indirectly alter beneficial microbial symbionts. Cross-ocean bacterial colonization of polyvinylchloride (PVC) is dominated by *Rhodobacterales* (*22*), a group of potentially opportunistic pathogens associated with outbreaks of several coral diseases (*23*). Additionally, recent studies have shown that experimental shading and low-light microenvironments can lead to anoxic conditions favoring the formation of polymicrobial mats characteristic of black band disease (*24*).

By disproportionately reducing the composition or abundance of structurally complex reef-building coral species through disease, widespread distribution of plastic waste may have negative consequences for biodiversity and people (*25*). For example, on coral reefs, the loss of structural habitat availability for reef organisms has been shown to reduce fishery productivity by a factor of 3 (*26*).

Climate-related disease outbreaks have already affected coral reefs globally and are projected to increase in frequency and severity as ocean temperatures rise (*27*). With more than 275 million people relying on coral reefs for food, coastal protection, tourism income, and cultural importance (*2*), moderating disease outbreak risks in the ocean will be vital for improving both human and ecosystem health. Our study indicates that decreasing the levels of plastic debris entering the ocean by improving waste management infrastructure is critical for reducing the amount of debris on coral reefs and the associated risk of disease and structural damage.

REFERENCES AND NOTES

1. C. D. Harvell *et al.*, *Science* **296**, 2158–2162 (2002).
2. L. Burke, K. Reytar, M. Spalding, A. Perry, *Reefs at Risk Revisted* (World Resources Institute, 2011).
3. I. V. Kirstein *et al.*, *Mar. Environ. Res.* **120**, 1–8 (2016).
4. S. Oberbeckmann, M. G. Löder, M. Labrenz, *Environ. Chem.* **12**, 551–562 (2015).
5. C. A. De Tender *et al.*, *Environ. Sci. Technol.* **49**, 9629–9638 (2015).
6. K. M. Morrow, R. Ritson-Williams, C. Ross, M. R. Liles, V. J. Paul, *PLOS ONE* **7**, e44859 (2012).
7. M. M. Nugues, G. W. Smith, R. J. Hooidonk, M. I. Seabra, R. P. Bak, *Ecol. Lett.* **7**, 919–923 (2004).
8. J. E. Smith *et al.*, *Ecol. Lett.* **9**, 835–845 (2006).
9. M. J. Sweet, J. C. Bythell, M. M. Nugues, *PLOS ONE* **8**, e69717 (2013).
10. E. R. Zettler, T. J. Mincer, L. A. Amaral-Zettler, *Environ. Sci. Technol.* **47**, 7137–7146 (2013).
11. D. Bourne, T. Ainsworth, F. Pollock, B. Willis, *Coral Reefs* **34**, 233–242 (2015).
12. J. R. Jambeck *et al.*, *Science* **347**, 768–771 (2015).
13. F. M. Fazey, P. G. Ryan, *Environ. Pollut.* **210**, 354–360 (2016).
14. D. K. Barnes, *Nature* **416**, 808–809 (2002).
15. Materials and methods are available as supplementary materials.
16. S. Altizer, R. S. Ostfeld, P. T. J. Johnson, S. Kutz, C. D. Harvell, *Science* **341**, 514–519 (2013).
17. J. B. Lamb, D. H. Williamson, G. R. Russ, B. L. Willis, *Ecology* **96**, 2555–2567 (2015).
18. L. E. Richardson, N. A. J. Graham, A. S. Hoey, *Sci. Rep.* **7**, 7557 (2017).
19. J. B. Lamb *et al.*, *Philos. Trans. R. Soc. London Ser. B* **371**, 20150210 (2016).
20. L. D. Mydlarz, L. E. Jones, C. D. Harvell, *Annu. Rev. Ecol. Evol. Syst.* **37**, 251–288 (2006).
21. C. A. Page, B. L. Willis, *Coral Reefs* **27**, 257–272 (2008).
22. H. Dang, T. Li, M. Chen, G. Huang, *Appl. Environ. Microbiol.* **74**, 52–60 (2008).
23. N. Soffer, J. Zaneveld, R. Vega Thurber, *Environ. Microbiol.* **17**, 1203–1218 (2015).
24. M. S. Glas, Y. Sato, K. E. Ulstrup, D. G. Bourne, *ISME J.* **6**, 1526–1534 (2012).
25. B. Worm *et al.*, *Science* **314**, 787–790 (2006).
26. A. Rogers, J. L. Blanchard, P. J. Mumby, *Curr. Biol.* **24**, 1000–1005 (2014).
27. J. Maynard *et al.*, *Nat. Clim. Chang.* **5**, 688–694 (2015).

ACKNOWLEDGMENTS

We thank S. Atto, S. Beveridge, F. J. Pollock, G. Torda, A. Tracy, A. Wenger, C. Wood, S. Yusuf, and S. Piromvaragorn for their assistance. This research was supported by The Nature Conservancy NatureNet Science Fellowship, an AIMS@JCU (Australian Institute of Marine Science at James Cook University) Postgraduate Scholarship, the NSF Ecology of Infectious Marine Disease Research Coordination Network [Division of Ocean Sciences (OCE) award 1215977], the Capturing Coral Reef and Ecosystem Related Services (CCRES) Project funded by the Global Environment Facility and the World Bank (project ID P123933), NOAA/National Ocean Service (NOS)–HIMB agreement code MOA-2009-039, Australian Research Council (grant CE0561435), and the Environmental Defense Fund Innovation for Impact partnership with the Atkinson Center for a Sustainable Future at Cornell University. Surveys in Australia were conducted under Great Barrier Reef Marine Park Authority permits G10/33393.1 and G12/35232.1. All data and code to understand and assess the conclusions of this research are available in the main text, supplementary materials, and via the Dryad Digital Repository (https://doi.org/10.5061/dryad.mp480).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6374/460/suppl/DC1
Materials and Methods
Figs. S1 and S2
Tables S1 to S14
References (*28*–*30*)

31 October 2017; accepted 21 December 2017
10.1126/science.aar3320

NEUROSCIENCE

Learning and attention reveal a general relationship between population activity and behavior

A. M. Ni, D. A. Ruff, J. J. Alberts, J. Symmonds, M. R. Cohen*

Prior studies have demonstrated that correlated variability changes with cognitive processes that improve perceptual performance. We tested whether correlated variability covaries with subjects' performance—whether performance improves quickly with attention or slowly with perceptual learning. We found a single, consistent relationship between correlated variability and behavioral performance, regardless of the time frame of correlated variability change. This correlated variability was oriented along the dimensions in population space used by the animal on a trial-by-trial basis to make decisions. That subjects' choices were predicted by specific dimensions that were aligned with the correlated variability axis clarifies long-standing paradoxes about the relationship between shared variability and behavior.

The responses of pairs of neurons to repeated presentations of the same stimulus are typically correlated [quantified as noise correlations, or spike count correlations (r_{SC})] (*1*, *2*). Prior electrophysiological studies have shown that these correlations change with cognitive processes that affect perceptual performance (*2*–*4*). However, theoretical work has suggested that this correlated variability may not affect the information encoded by a neuronal population in a manner that influences a subject's decisions (*5*, *6*).

We therefore measured the relationship between neuronal population activity and performance by studying two processes that both improve visual performance but on very different time scales: attention (*7*) and perceptual learning (*8*). By observing attention and learning in the same behavioral trials and neuronal populations, we identified the dimensions of population activity that matter most for behavior.

We recorded from neuronal populations in V4 (*3*, *4*, *7*–*9*) in two rhesus monkeys with chronically implanted microelectrode arrays (*3*). The monkeys detected changes in the orientation of either of two Gabor stimuli (Fig. 1A): one placed within the receptive fields (RFs) of the recorded neurons and one in the opposite hemifield (Fig. 1B). We measured attention effects within a single session and learning effects across sessions (Fig. 1C).

Attention and perceptual learning improved performance and affected neuronal population responses in similar ways (Fig. 2 and figs. S1 and S2). Both processes were associated with decreases in the mean-normalized trial-to-trial variance (Fano factor) of individual units and the correlated variability between pairs of units (Fig. 2, C, D, J, and K) in response to repeated presentations of the same stimulus (figs. S3 and S4). These variability changes occurred only in the context of the task (variability measured during passive fixation was constant throughout training) (Fig. 2, F, G, M, and N).

Recent theoretical work suggests that only correlated variability along the dimensions in neuronal population space that encode task-relevant stimulus information can limit information coding (*5*, *6*). Determining whether correlated variability lies along these dimensions is experimentally unfeasible because it would require recordings from a very large number of neurons over an even larger number of trials.

Instead, we assessed the importance of attention- and perceptual learning–related changes

Department of Neuroscience and Center for the Neural Basis of Cognition, University of Pittsburgh, Pittsburgh, PA 15260, USA.
*Corresponding author. Email: cohenm@pitt.edu

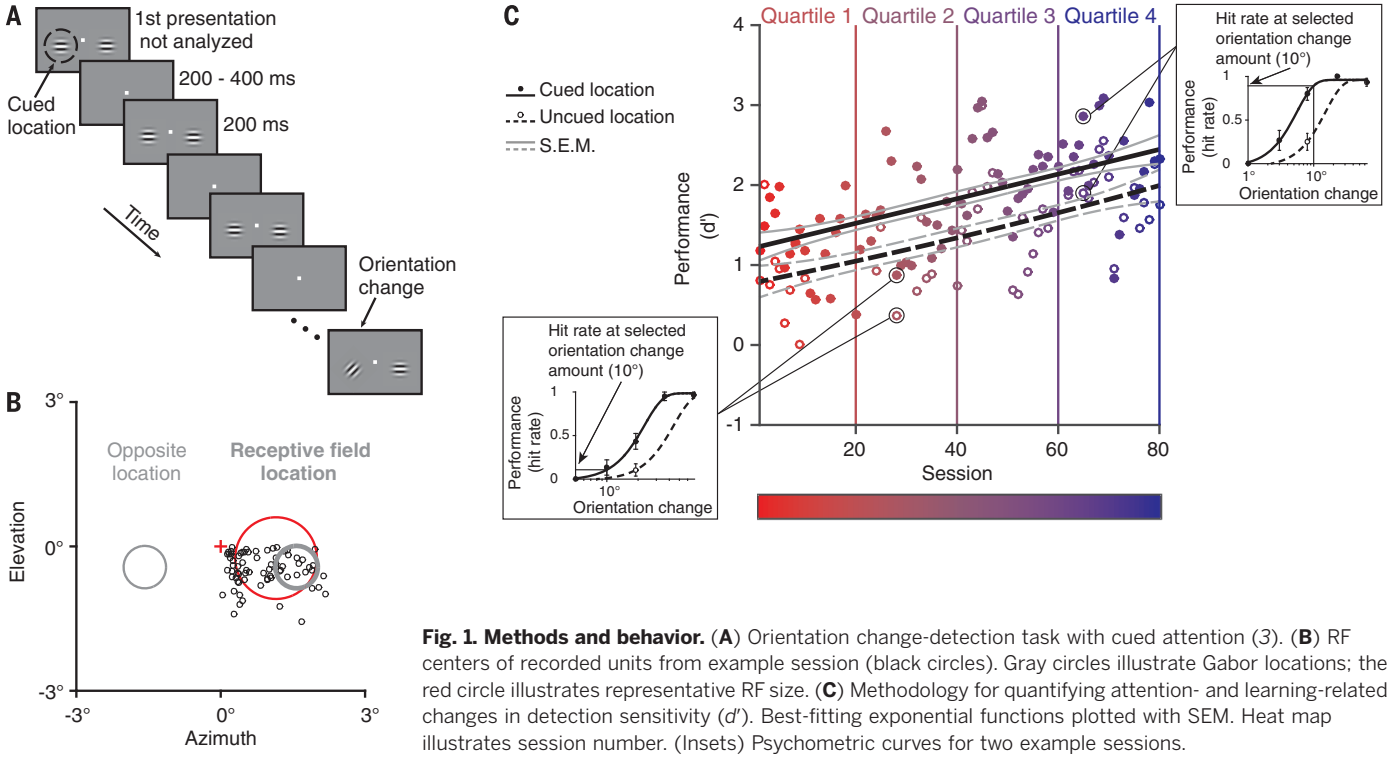


Fig. 1. Methods and behavior. (**A**) Orientation change-detection task with cued attention (*3*). (**B**) RF centers of recorded units from example session (black circles). Gray circles illustrate Gabor locations; the red circle illustrates representative RF size. (**C**) Methodology for quantifying attention- and learning-related changes in detection sensitivity (*d'*). Best-fitting exponential functions plotted with SEM. Heat map illustrates session number. (Insets) Psychometric curves for two example sessions.

in correlated variability by investigating their relationship to behavior. There was a single, robust relationship between correlated variability and perceptual performance, whether changes in performance happened quickly (attention) (Fig. 3, A and B) or slowly (learning) (Fig. 3, C and D). This relationship was robust even when we removed the main effects of attention and learning (Fig. 3, E and F).

We analyzed the responses of V1 neurons (*7*, *8*) in animals performing the same attention task. Unlike in V4, correlated variability in V1 was not correlated with performance (fig. S5).

Both attention and perceptual learning improved the performance of a cross-validated, optimal linear stimulus decoder (fig. S6). However, the relationship between correlated variability in V4 and performance (Fig. 3) seems at odds with theoretical work that suggests most correlated variability should not affect the stimulus information that can be gleaned from an optimal decoder (*6*).

To examine the relationship between correlated variability and performance more directly, we developed a single-trial measure of correlated variability. We performed principal component analysis (PCA) on population responses to the same repeated stimuli used to compute spike count correlations (fig. S3), meaning that the first PC is by definition the axis that explains more of the correlated variability than any other dimension (Fig. 4, A and B, x axis). Consistent with the recent observation that correlated variability is typically low dimensional (*10*–*12*), the variance explained by the first PC was strongly related to the magnitude of correlated variability in each session, even when we accounted for the changes caused by attention and learning (Fig. 4, C and D, and fig. S7) and trial-averaged firing rates (figs. S8 to S11). Like correlated variability (Fig. 3), the proportion of variance explained by the first PC was correlated with behavioral performance (d') across all sessions [Monkey 1, correlation coefficient (R) = −0.42, $P < 10^{-13}$; Monkey 2, R = −0.62, $P < 10^{-15}$].

These analyses show that projection on this first PC is a suitable proxy for pairwise spike count correlations. We used this measure to assess the importance of correlated variability to the monkey by determining whether population activity along this first PC can predict the monkey's choices on a trial-by-trial basis.

Activity along this first PC (and therefore correlated variability) had a much stronger relationship with the monkey's behavior than it would if the monkey used an optimal stimulus decoder. A linear, cross-validated choice decoder (Fig. 4A) could detect differences in hit versus miss trial responses to the changed stimulus from V4 population activity along the first PC alone as well as it could from our full data set (Fig. 4, E and F, and fig. S12). By contrast, although the performance of the stimulus decoder (Fig. 4A) at detecting differences in V4 neuronal population responses to the previous stimulus (the stimulus before the change) versus the changed stimulus was unsurprisingly better overall than

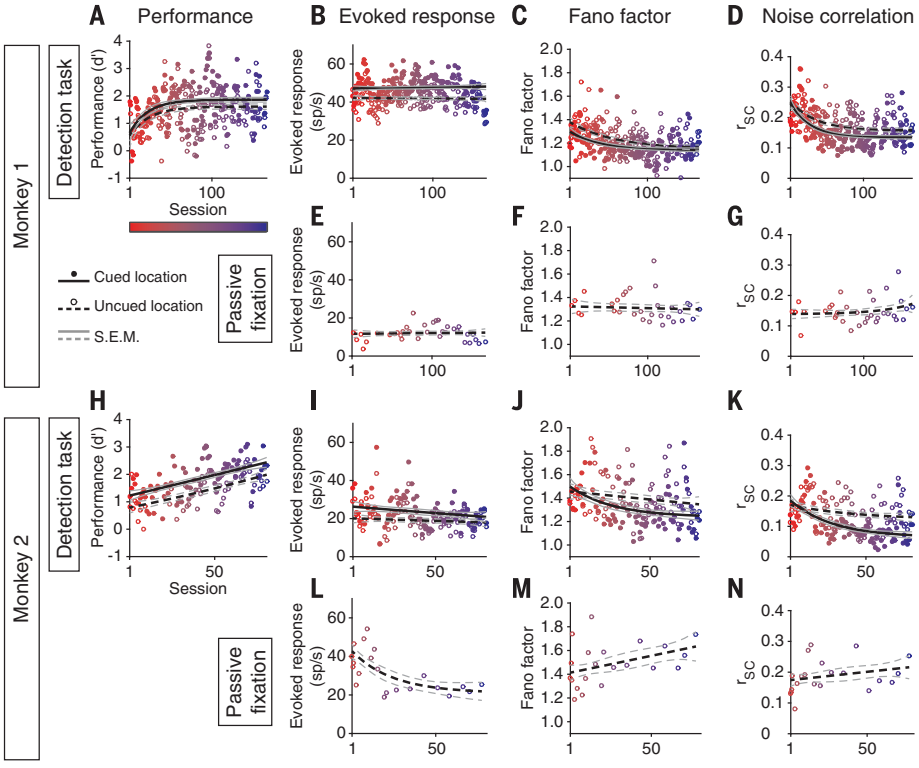


Fig. 2. Summary of behavioral and neuronal effects of attention and perceptual learning. All changes were significantly different than 0 except where indicated (t tests; $P < 10^{-3}$). Conventions are as in Fig. 1C. (**A** and **H**) Sensitivity (d') increased with both attention and learning. (**B**, **E**, **I**, and **L**) Evoked response (firing – baseline rate) increased with attention but did not change consistently with learning or passive fixation ($P > 0.05$). (**C**, **D**, **J**, and **K**) Fano factor and correlated variability decreased with attention and learning, but (**F**, **G**, **M**, and **N**) not during passive fixation ($P > 0.05$).

the performance of the choice decoder (Fig. 4, E and F, insets), the relative influence of the first PC was weaker. The performance of the stimulus decoder was much worse when based on the first PC alone versus our full data set (Fig. 4, E and F).

It is difficult to determine from extracellular recordings whether choice-predictive signals come from a bottom-up, causal relationship between sensory responses and decisions or from trial-to-trial variability from cognitive factors or post-decision signals (*13*). A recent study identifying the directionality of choice-predictive signals in mouse sensory cortex found that they are both bottom-up and top-down in origin (*14*). However, the time course of the choice-predictive activity in our data suggests that it occurs before the decision is made. We based our choice decoder on the first 70 ms of the evoked responses (after accounting for the response latency of V4 neurons). Choice-predictive activity was as strong in the first half of this time frame (60 to 95 ms) as in the second half (96 to 130 ms; paired t test per monkey, $P > 0.05$). That the choice-predictive activity described here was present during the full decision-making period suggests that it did not reflect post-decision feedback.

Our results, combined with functional imaging in humans (*8*) and other multielectrode recording studies (*15*, *16*), suggest that learning is best studied by focusing on populations of neurons. Functional imaging studies, which use measures that are related to the activity of large neural populations, find consistent learning-related changes in both V1 and V4 (*8*, *17*), as opposed to single-unit studies (*8*). Similarly, attention studies suggest that changes in population sensitivity are largely explained by cross-neuron correlations as opposed to single-neuron effects (*3*, *4*).

The robust relationship between correlated variability and perceptual performance suggests that although attention and learning mechanisms act on different time scales (fig. S13), they share a common computation. Some characteristics of this computation are informed by recent studies showing that changes in a low rank modulator can account for the attention-related changes in rate, Fano factor, and correlated variability (*11*, *12*). Attention and learning may decrease the strength of such a modulator by changing the balance of inhibition and excitation (*10*), which may improve information coding and the information that is communicated downstream (*18*).

Our most puzzling finding is that the attention- and learning-related changes in average noise correlation were so closely linked to performance but would likely have a minimal effect on performance if the monkeys read out visual information optimally. Similarly, a prior

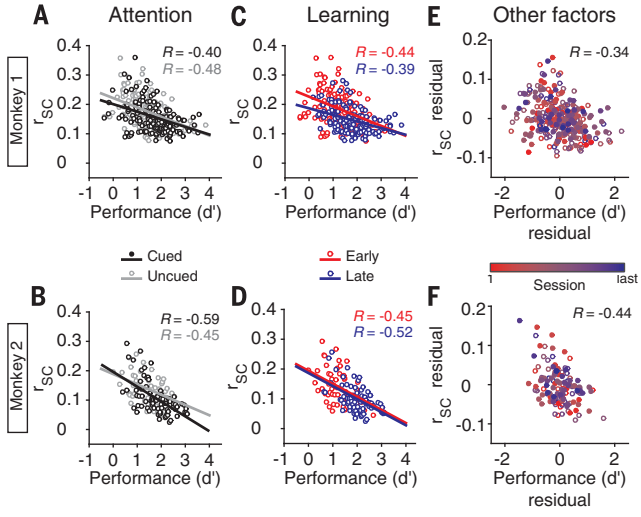


Fig. 3. The relationship between correlated variability and performance is the same for attention and perceptual learning. Mean r_{SC} and d' were significantly correlated across sessions ($P < 10^{-3}$). (**A** and **B**) Relationship between r_{SC} and d' was indistinguishable between attention conditions (Fisher *z* Pearson-Filon tests; $P > 0.05$). (**C** and **D**) Relationship between r_{SC} and d' was indistinguishable for the first versus second half of learning ($P > 0.05$). (**E** and **F**) Relationships persisted after removing attention and learning effects (residuals of exponential fits in Fig. 2; $P < 10^{-3}$; analyses of variance, $P > 0.05$).

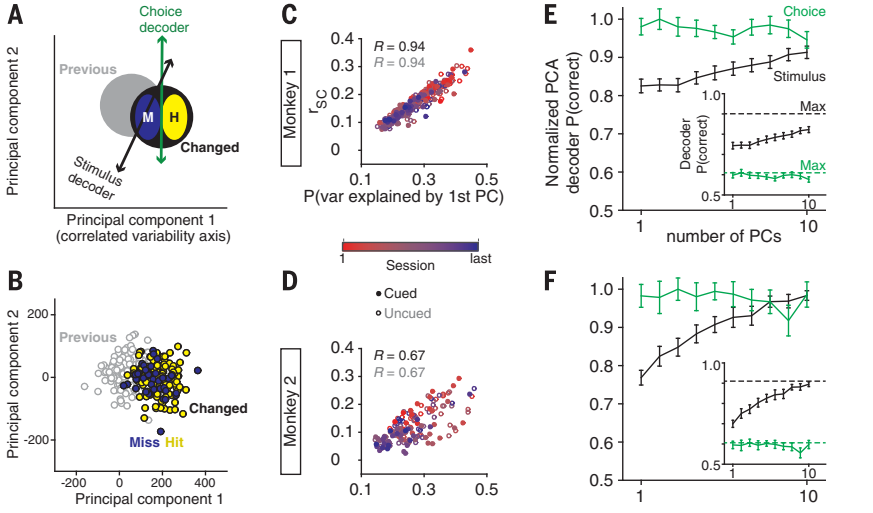


Fig. 4. Correlated variability is more closely aligned with choices than would be expected from an optimal stimulus decoder. (**A**) Schematic showing how we obtained our single-trial measure of correlated variability. We performed PCA on responses to the stimulus before the change (gray) and to the changed stimulus (black). The stimulus decoder detects differences between the responses to the previous and changed stimuli. The choice decoder detects differences between responses to misses and hits. (**B**) Example data set showing that the animal's choices are more aligned with the first PC (*x* axis) than the difference between the previous and changed stimuli, which depends on both the first and second PC. (**C** and **D**) Mean r_{SC} is highly correlated with proportion of variance explained by the first PC ($P < 10^{-11}$; residuals, $P < 10^{-9}$). (**E** and **F**) Performance of the choice and stimulus decoders normalized to their maximum performance (with the full data set), with SEM. (Inset) Raw decoder performance.

study found that correlations depend on training experience but did not find a relationship between shared variability and information coding (*19*). Correlated variability should only affect the performance of an optimal decoder when it is aligned with the stimulus dimension being decoded (*6*). Therefore, the relationship between correlated variability and performance suggests that our monkeys performed suboptimally.

We thus hypothesize that sensory information is decoded in a way that is optimal for the large number of stimuli and tasks that the animals encounter in their natural environment rather than the particular set of stimuli in our task. Traditionally, optimal decoders are trained to discriminate a particular set of stimuli that vary only in one stimulus dimension. This scenario implies a two-step decision process: identifying the stimulus (to optimize the decoder) and then decoding it. If animals could successfully identify the stimulus, they would perform perfectly on our change-detection task.

Instead, animals may use a more general decoder that could, for example, identify the orientation of any stimulus in any task, meaning that optimal weights would be tuned and noise correlations related to all stimulus features for which the neurons are selective. Noise correlations depend on tuning similarity for all stimulus features (*6*). Therefore, correlated variability is likely aligned with the dimension that is decoded by a general decoder, meaning that noise correlation decreases would improve performance. Several of the studies that suggest monkeys do behave optimally are those that used multisensory stimuli (*20*). Determining whether there is evidence that monkeys use decoders that are optimized for diverse stimuli and tasks will be an important avenue for future work. Our results suggest that the relationship between behavior and population activity is a powerful tool for understanding neural computation.

REFERENCES AND NOTES

1. E. Zohary, M. N. Shadlen, W. T. Newsome, *Nature* **370**, 140–143 (1994).
2. M. R. Cohen, A. Kohn, *Nat. Neurosci.* **14**, 811–819 (2011).
3. M. R. Cohen, J. H. R. Maunsell, *Nat. Neurosci.* **12**, 1594–1600 (2009).
4. J. F. Mitchell, K. A. Sundberg, J. H. Reynolds, *Neuron* **63**, 879–888 (2009).
5. R. Moreno-Bote *et al.*, *Nat. Neurosci.* **17**, 1410–1417 (2014).
6. A. Kohn, R. Coen-Cagli, I. Kanitscheider, A. Pouget, *Annu. Rev. Neurosci.* **39**, 237–256 (2016).
7. J. H. R. Maunsell, *Annu. Rev. Vis. Sci.* **1**, 373–391 (2015).
8. T. Watanabe, Y. Sasaki, *Annu. Rev. Psychol.* **66**, 197–221 (2015).
9. D. A. Ruff, M. R. Cohen, *Nat. Neurosci.* **17**, 1591–1597 (2014).
10. T. Kanashiro, G. K. Ocker, M. R. Cohen, B. Doiron, *eLife* **6**, e23978 (2017).
11. N. C. Rabinowitz, R. L. Goris, M. Cohen, E. P. Simoncelli, *eLife* **4**, e08998 (2015).
12. A. S. Ecker, G. H. Denfield, M. Bethge, A. S. Tolias, *J. Neurosci.* **36**, 1775–1789 (2016).
13. H. Nienborg, B. G. Cumming, *Nature* **459**, 89–92 (2009).
14. S. E. Kwon, H. Yang, G. Minamisawa, D. H. O'Connor, *Nat. Neurosci.* **19**, 1243–1249 (2016).
15. J. M. Jeanne, T. O. Sharpee, T. Q. Gentner, *Neuron* **78**, 352–363 (2013).
16. Y. Yan *et al.*, *Nat. Neurosci.* **17**, 1380–1387 (2014).
17. J. F. Jehee, S. Ling, J. D. Swisher, R. S. van Bergen, F. Tong, *J. Neurosci.* **32**, 16747–53a (2012).
18. D. A. Ruff, M. R. Cohen, *J. Neurosci.* **36**, 7523–7534 (2016).
19. Y. Gu *et al.*, *Neuron* **71**, 750–761 (2011).
20. C. Chandrasekaran, *Curr. Opin. Neurobiol.* **43**, 25–34 (2017).

ACKNOWLEDGMENTS

M.R.C. is supported by U.S. NIH grants 4R00EY020844-03, R01 EY022930, and Core Grant P30 EY008098s; a Whitehall Foundation grant; a Klingenstein-Simons Fellowship; a Sloan Research Fellowship; a McKnight Scholar Award; and a grant from the Simons Foundation. A.M.N. is supported by a fellowship from the Simons Foundation. We thank K. McCracken for technical assistance and J. H. R. Maunsell and A. Kohn for comments on a previous version of this manuscript. A.M.N., D.A.R., and M.R.C. designed the experiments; A.M.N., D.A.R., J.J.A., and J.S. collected the data; A.M.N. performed the analyses; and A.M.N. and M.R.C. wrote the paper. The authors declare no competing financial interests. Data analyzed in this manuscript are available at https://pitt.box.com/v/NiRuffAlbertsSymmondsCohen2017.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6374/463/suppl/DC1
Materials and Methods
Figs. S1 to S13
References (*21*, *22*)

7 June 2017; resubmitted 21 August 2017
Accepted 21 December 2017
10.1126/science.aao0284

HUMAN IMPACTS

Moving in the Anthropocene: Global reductions in terrestrial mammalian movements

Marlee A. Tucker,[1,2]* Katrin Böhning-Gaese,[1,2] William F. Fagan,[3,4] John M. Fryxell,[5] Bram Van Moorter,[6] Susan C. Alberts,[7] Abdullahi H. Ali,[8] Andrew M. Allen,[9,10] Nina Attias,[11] Tal Avgar,[12] Hattie Bartlam-Brooks,[13] Buuveibaatar Bayarbaatar,[14] Jerrold L. Belant,[15] Alessandra Bertassoni,[16] Dean Beyer,[17] Laura Bidner,[18] Floris M. van Beest,[19] Stephen Blake,[20,21] Niels Blaum,[22] Chloe Bracis,[1,2] Danielle Brown,[23] P. J. Nico de Bruyn,[24] Francesca Cagnacci,[25,26] Justin M. Calabrese,[3,27] Constança Camilo-Alves,[28,29] Simon Chamaillé-Jammes,[30] Andre Chiaradia,[31,32] Sarah C. Davidson,[33,20] Todd Dennis,[34] Stephen DeStefano,[35] Duane Diefenbach,[36] Iain Douglas-Hamilton,[37,38] Julian Fennessy,[39] Claudia Fichtel,[40] Wolfgang Fiedler,[20] Christina Fischer,[41] Ilya Fischhoff,[42] Christen H. Fleming,[3,27] Adam T. Ford,[43] Susanne A. Fritz,[1,2] Benedikt Gehr,[44] Jacob R. Goheen,[45] Eliezer Gurarie,[3,46] Mark Hebblewhite,[47] Marco Heurich,[48,49] A. J. Mark Hewison,[50] Christian Hof,[1] Edward Hurme,[3] Lynne A. Isbell,[18,51] René Janssen,[52] Florian Jeltsch,[22] Petra Kaczensky,[6,53] Adam Kane,[54] Peter M. Kappeler,[40] Matthew Kauffman,[55] Roland Kays,[56,57] Duncan Kimuyu,[58] Flavia Koch,[40,59] Bart Kranstauber,[44] Scott LaPoint,[20,60] Peter Leimgruber,[27] John D. C. Linnell,[6] Pascual López-López,[61] A. Catherine Markham,[62] Jenny Mattisson,[6] Emilia Patricia Medici,[63,64] Ugo Mellone,[65] Evelyn Merrill,[12] Guilherme de Miranda Mourão,[66] Ronaldo G. Morato,[67] Nicolas Morellet,[50] Thomas A. Morrison,[68] Samuel L. Díaz-Muñoz,[69,70] Atle Mysterud,[71] Dejid Nandintsetseg,[1,2] Ran Nathan,[72] Aidin Niamir,[1] John Odden,[73] Robert B. O'Hara,[1,74] Luiz Gustavo R. Oliveira-Santos,[75] Kirk A. Olson,[14] Bruce D. Patterson,[76] Rogerio Cunha de Paula,[67] Luca Pedrotti,[77] Björn Reineking,[78,79] Martin Rimmler,[80] Tracey L. Rogers,[81] Christer Moe Rolandsen,[6] Christopher S. Rosenberry,[82] Daniel I. Rubenstein,[83] Kamran Safi,[20,84] Sonia Saïd,[85] Nir Sapir,[86] Hall Sawyer,[87] Niels Martin Schmidt,[19,88] Nuria Selva,[89] Agnieszka Sergiel,[89] Enkhtuvshin Shiilegdamba,[14] João Paulo Silva,[90,91,92] Navinder Singh,[9] Erling J. Solberg,[6] Orr Spiegel,[93] Olav Strand,[6] Siva Sundaresan,[94] Wiebke Ullmann,[22] Ulrich Voigt,[95] Jake Wall,[37] David Wattles,[35] Martin Wikelski,[20,84] Christopher C. Wilmers,[96] John W. Wilson,[97] George Wittemyer,[37,98] Filip Zięba,[99] Tomasz Zwijacz-Kozica,[99] Thomas Mueller[1,2,27]*

Animal movement is fundamental for ecosystem functioning and species survival, yet the effects of the anthropogenic footprint on animal movements have not been estimated across species. Using a unique GPS-tracking database of 803 individuals across 57 species, we found that movements of mammals in areas with a comparatively high human footprint were on average one-half to one-third the extent of their movements in areas with a low human footprint. We attribute this reduction to behavioral changes of individual animals and to the exclusion of species with long-range movements from areas with higher human impact. Global loss of vagility alters a key ecological trait of animals that affects not only population persistence but also ecosystem processes such as predator-prey interactions, nutrient cycling, and disease transmission.

With approximately 50 to 70% of Earth's land surface currently modified for human activities (*1*), patterns of biodiversity and ecosystem functions worldwide are changing (*2*). The expanding footprint of human activities not only is causing the loss of habitat and biodiversity, but is also affecting how animals move through fragmented and disturbed habitats. The extent to which animal movements are affected by anthropogenic effects on the structure and composition of landscapes and resource changes has been explored only in local geographic regions or within single species. Such studies typically report decreasing animal movements—for example, as a result of habitat fragmentation, barrier effects, or resource changes (*3*–*6*)—with only a few studies reporting longer movements as a result of habitat loss or altered migration routes (*7*, *8*).

We conducted a global comparative study of how the human footprint affects movements of terrestrial nonvolant mammals, using Global Positioning System (GPS) location data of 803 individuals from 57 mammal species (Fig. 1 and table S2). Mean species mass ranged from 0.49 to 3940 kg and included herbivores, carnivores, and omnivores (n = 28, 11, and 18 species, respectively). For each individual, we annotated locations with the Human Footprint Index (HFI), an index with a global extent that combines multiple proxies of human influence: the extent of built environments, crop land, pasture land, human population density, nighttime lights, railways, roads, and navigable waterways (*9*) (see supplementary materials for details). The HFI ranges from 0 (natural environments: e.g., the Brazilian Pantanal) to 50 (high-density built environments: e.g., New York City).

In addition to the human footprint, we included other covariates that are known to influence mammalian movements. Because individuals may need to cover a larger area to gather sufficient resources, mammals generally move farther in environments with lower productivity (*10*). To capture this effect, we annotated locations with the Normalized Difference Vegetation Index (NDVI), a well-established, satellite-derived measure of resource abundance for both herbivores and carnivores (*11*). Because an allometric scaling relationship shows that animals of greater body size usually move farther (*12*), and because diet may influence movements as a result of differences in foraging costs and availability of resource types (*13*, *14*), we annotated the database with species averages for body size and dietary guild (i.e., carnivore, herbivore, or omnivore).

We then calculated displacements as the distance between subsequent GPS locations of each individual at nine time scales (*15*) ranging from 1 hour to 10 days. For each individual at each time scale, we calculated the 0.5 and 0.95 quantile of displacement. The combination of different time scales and quantile allowed us to examine the effect of the human footprint on both the median (0.5 quantile) and long-distance (0.95 quantile) movements for within-day movements (e.g., 1-hour time scale) up to longer time displacements of more than 1 week (e.g., 10-day time scale). We used linear mixed-effects models that, in addition to all covariates (i.e., NDVI, body mass, diet), also accounted for taxonomy and spatial autocorrelation (see supplementary materials for details).

We found strong negative effects of the human footprint on median and long-distance displacements of terrestrial mammals (Fig. 2, Fig. 3A, and table S3). Displacements of individuals (across species) living in areas of high footprint (HFI = 36) were shorter than displacements of individuals living in areas of low footprint (HFI = 0) by as much as a factor of 3. For example, median displacements for carnivores over 10 days were 3.3 ± 1.4 km (SE) in areas of high footprint versus 6.9 ± 1.3 km in areas of low footprint (Fig. 2A and table S3). Likewise, the maximum displacement distances for carnivores at the 10-day scale averaged 6.6 ± 1.4 km in areas of high footprint versus 21.5 ± 1.4 km in areas of low footprint (Fig. 2B and table S3). The effect was significant on all temporal scales with 8 hours or more between locations.

The effect was not significant at shorter time scales (Fig. 3A, 1 to 4 hours), which suggests that the human footprint affects ranging behavior and area use over longer time scales, rather than altering individual travel speeds (i.e., individuals may travel at the same speed if measured across short time intervals, but have more tortuous movements in areas of higher human footprint and thus remain in the same locale if displacement is measured across longer time intervals).

Reduction in movement may be attributable to (i) an individual-behavioral effect, where individuals alter their movements relative to the human footprint, or (ii) a species occurrence effect, where certain species that exhibit long-range movement simply do not occur in areas of high human footprint. To disentangle these two effects, we ran additional models where we separated the HFI into two components: (i) the individual-behavioral effect represented by the individual variability of HFI relative to the species mean (i.e., the individual HFI minus the species mean HFI), and (ii) the species occurrence effect as the mean HFI for each species. Results from the two-component model indicate behavioral as well as species effects. We found a significant behavioral effect on median displacements and on long-distance displacements (0.95 quantile) at most time scales (from 8 hours to 10 days) (fig. S2A and table S4). The species occurrence effect was significant only over longer time scales (128- and 256-hour periods, or 5 and 10 days, respectively) (fig. S2B and table S4). However, we note that the estimate of the species occurrence effect is conservative because our model incorporated taxonomy as a random effect. Some variability in the data may have been accounted for by the species-level random effect rather than the species-level HFI (see table S3).

In addition to the human footprint effect, body mass, dietary guild, and resource availability were also related to movement distances. First, as expected from allometric scaling and established relationships of body size with home range size (*14*) and migration distance (*16*), larger species traveled farther than smaller species (Fig. 3C and tables S3 and S4). Second, we found a negative

[1]Senckenberg Biodiversity and Climate Research Centre, Senckenberg Gesellschaft für Naturforschung, 60325 Frankfurt (Main), Germany. [2]Department of Biological Sciences, Goethe University, 60438 Frankfurt (Main), Germany. [3]Department of Biology, University of Maryland, College Park, MD 20742, USA. [4]SESYNC, University of Maryland, Annapolis, MD 21401, USA. [5]Department of Integrative Biology, University of Guelph, Guelph, Ontario N1G 2W1, Canada. [6]Norwegian Institute for Nature Research, P.O. Box 5685 Torgard, NO-7485 Trondheim, Norway. [7]Departments of Biology and Evolutionary Anthropology, Duke University, Durham, NC 27708, USA. [8]Hirola Conservation Programme, Garissa, Kenya. [9]Department of Wildlife, Fish and Environmental Studies, Swedish University of Agricultural Sciences, Umeå 90183, Sweden. [10]Institute for Water and Wetland Research, Department of Animal Ecology and Physiology, Radboud University, 6500GL Nijmegen, Netherlands. [11]Ecology and Conservation Graduate Program, Federal University of Mato Grosso do Sul, Campo Grande, MS, Brazil. [12]Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada. [13]Structure and Motion Laboratory, Royal Veterinary College, University of London, London NW1 0TU, UK. [14]Wildlife Conservation Society, Mongolia Program, Ulaanbaatar, Mongolia. [15]Carnivore Ecology Laboratory, Forest and Wildlife Research Center, Mississippi State University, Box 9690, Mississippi State, MS, USA. [16]Animal Biology Postgraduate Program, São Paulo State University, São José do Rio Preto, SP 15054-000, Brazil. [17]Michigan Department of Natural Resources, 1990 U.S. 41 South, Marquette, MI 49855, USA. [18]Department of Anthropology, University of California, Davis, CA 95616, USA. [19]Department of Bioscience, Aarhus University, 4000 Roskilde, Denmark. [20]Max Planck Institute for Ornithology, Vogelwarte Radolfzell, D-78315 Radolfzell, Germany. [21]Wildlife Conservation Society, Bronx, NY 10460, USA. [22]University of Potsdam, Plant Ecology and Nature Conservation, 14476 Potsdam, Germany. [23]Department of Biology, Middle Tennessee State University, Murfreesboro, TN 37132, USA. [24]Mammal Research Institute, Department of Zoology and Entomology, University of Pretoria, Hatfield 0028, Gauteng, South Africa. [25]Department of Biodiversity and Molecular Ecology, Research and Innovation Centre, Fondazione Edmund Mach, 38010 San Michele all'Adige (TN), Italy. [26]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. [27]Smithsonian Conservation Biology Institute, National Zoological Park, Front Royal, VA, USA. [28]Departamento de Fitotecnia, Universidade de Évora, Pólo da Mitra, 7002-554 Évora, Portugal. [29]ICAAM–Institute of Mediterranean Agricultural and Environmental Sciences, University of Évora, Évora, Portugal. [30]Centre d'Ecologie Fonctionnelle et Evolutive UMR 5175, CNRS–Université de Montpellier–Université Paul-Valéry Montpellier–EPHE, 34293 Montpellier Cedex 5, France. [31]Phillip Island Nature Parks, Victoria, Australia. [32]School of Biological Sciences, Monash University, Melbourne, Australia. [33]Department of Civil, Environmental and Geodetic Engineering, Ohio State University, Columbus, OH 43210, USA. [34]Department of Biology, Fiji National University, P.O. Box 5529, Natabua, Lautoka, Fiji Islands. [35]U.S. Geological Survey, Massachusetts Cooperative Fish and Wildlife Research Unit, University of Massachusetts, Amherst, MA 01003, USA. [36]U.S. Geological Survey, Pennsylvania Cooperative Fish and Wildlife Research Unit, Pennsylvania State University, University Park, PA 16802, USA. [37]Save the Elephants, P.O. Box 54667, Nairobi 00200, Kenya. [38]Department of Zoology, University of Oxford, Oxford OX1 3PS, UK. [39]Giraffe Conservation Foundation, P.O. Box 86099, Eros, Namibia. [40]German Primate Center, Behavioral Ecology and Sociobiology Unit, 37077 Göttingen, Germany. [41]Restoration Ecology, Department of Ecology and Ecosystem Management, Technische Universität München, 85354 Freising, Germany. [42]Cary Institute of Ecosystem Studies, Millbrook, NY 12545, USA. [43]Irving K. Barber School of Arts and Sciences, Unit 2: Biology, University of British Columbia, Okanagan Campus, Kelowna, BC V1V 1V7, Canada. [44]Department of Evolutionary Biology and Environmental Studies, University of Zurich, 8057 Zurich, Switzerland. [45]Department of Zoology and Physiology, University of Wyoming, Laramie, WY 82071, USA. [46]School of Environmental and Forest Sciences, University of Washington, Seattle, WA 98195, USA. [47]Wildlife Biology Program, Department of Ecosystem and Conservation Sciences, College of Forestry and Conservation, University of Montana, Missoula, MT 59812, USA. [48]Bavarian Forest National Park, Department of Conservation and Research, 94481 Grafenau, Germany. [49]Chair of Wildlife Ecology and Management, Albert Ludwigs University of Freiburg, 79106 Freiburg, Germany. [50]CEFS, Université de Toulouse, INRA, Castanet Tolosan, France. [51]Animal Behavior Graduate Group, University of California, Davis, CA 95616, USA. [52]Bionet Natuuronderzoek, 6171EL Stein, Netherlands. [53]Research Institute of Wildlife Ecology, University of Veterinary Medicine Vienna, A-1160 Vienna, Austria. [54]School of Biological, Earth and Environmental Sciences, University College Cork, Cork, Ireland. [55]U.S. Geological Survey, Wyoming Cooperative Fish and Wildlife Research Unit, Department of Zoology and Physiology, University of Wyoming, Laramie, WY, USA. [56]North Carolina Museum of Natural Sciences, Raleigh, NC 27601, USA. [57]Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC 27695, USA. [58]Department of Natural Resource Management, Karatina University, P.O. Box 1957-10101, Karatina, Kenya. [59]Department of Psychology, University of Lethbridge, Lethbridge, Alberta T1K 3M4, Canada. [60]Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY 10964, USA. [61]Cavanilles Institute of Biodiversity and Evolutionary Biology, Terrestrial Vertebrates Group, University of Valencia, E-46980 Paterna, Valencia, Spain. [62]Department of Anthropology, Stony Brook University, Stony Brook, NY 11794, USA. [63]International Union for Conservation of Nature (IUCN) Species Survival Commission (SSC) Tapir Specialist Group (TSG), Rua Licuala, 622, Damha 1, Campo Grande, CEP: 79046-150, Mato Grosso do Sul, Brazil. [64]IPÊ (Instituto de Pesquisas Ecológicas; Institute for Ecological Research), Caixa Postal 47, Nazaré Paulista, CEP: 12960-000, São Paulo, Brazil. [65]Vertebrates Zoology Research Group, Departamento de Ciencias Ambientales y Recursos Naturales, University of Alicante, Alicante, Spain. [66]Embrapa Pantanal, Corumbá, MS 79320-900, Brazil. [67]National Research Center for Carnivores Conservation, Chico Mendes Institute for the Conservation of Biodiversity, Atibaia-SP 12952-011, Brazil. [68]Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, UK. [69]Center for Genomics and Systems Biology, Department of Biology, New York University, New York, NY 10003, USA. [70]Department of Microbiology and Molecular Genetics, University of California, Davis, CA 95616, USA. [71]Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, Blindern, NO-0316 Oslo, Norway. [72]Movement Ecology Laboratory, Department of Ecology, Evolution and Behavior, Alexander Silberman Institute of Life Sciences, Hebrew University of Jerusalem, Jerusalem 91904, Israel. [73]Norwegian Institute for Nature Research, NO-0349 Oslo, Norway. [74]Department of Mathematical Sciences and Centre for Biodiversity Dynamics, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway. [75]Department of Ecology, Federal University of Mato Grosso do Sul, Campo Grande, MS 79070-900, Brazil. [76]Integrative Research Center, Field Museum of Natural History, Chicago, IL 60605, USA. [77]Consorzio Parco Nazionale dello Stelvio, Bormio (Sondrio), Italy. [78]Univ. Grenoble Alpes, Irstea, UR LESSEM, BP 76, 38402 St-Martin-d'Hères, France. [79]University of Bayreuth, BayCEER, 95447 Bayreuth, Germany. [80]Nationalpark Schwarzwald, 77889 Seebach, Germany. [81]Evolution and Ecology Research Centre and School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia. [82]Pennsylvania Game Commission, Harrisburg, PA 17110, USA. [83]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA. [84]Department of Biology, University of Konstanz, 78467 Konstanz, Germany. [85]Directorate of Studies and Expertise (DRE), Office National de la Chasse et de la Faune Sauvage, Montfort, 01330 Birieux, France. [86]Department of Evolutionary and Environmental Biology, University of Haifa, 3498838 Haifa, Israel. [87]Western Ecosystems Technology Inc., Laramie, WY 82070, USA. [88]Arctic Research Centre, Aarhus University, 8000 Aarhus C, Denmark. [89]Institute of Nature Conservation Polish Academy of Sciences, 31-120 Krakow, Poland. [90]REN Biodiversity Chair, CIBIO/InBIO Associate Laboratory, Universidade do Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal. [91]Centre for Applied Ecology "Prof. Baeta Neves"/InBIO Associate Laboratory, Instituto Superior de Agronomia, Universidade de Lisboa, Tapada da Ajuda, 1349-017 Lisbon, Portugal. [92]Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisbon, Portugal. [93]Department of Environmental Science and Policy, University of California, Davis, CA, USA. [94]Jackson Hole Conservation Alliance, Jackson, WY 83001, USA. [95]Institute for Terrestrial and Aquatic Wildlife Research, University of Veterinary Medicine Hannover-Foundation, 30173 Hannover, Germany. [96]Center for Integrated Spatial Research, Environmental Studies Department, University of California, Santa Cruz, CA 95060, USA. [97]Department of Zoology and Entomology, University of Pretoria, Hatfield 0028, South Africa. [98]Department of Fish, Wildlife and Conservation Biology, Colorado State University, Fort Collins, CO 80523, USA. [99]Tatra National Park, 34-500 Zakopane, Poland.
***Corresponding author. Email: tucker.marlee@gmail.com (M.A.T.); thomas.mueller@senckenberg.de (T.M.)**

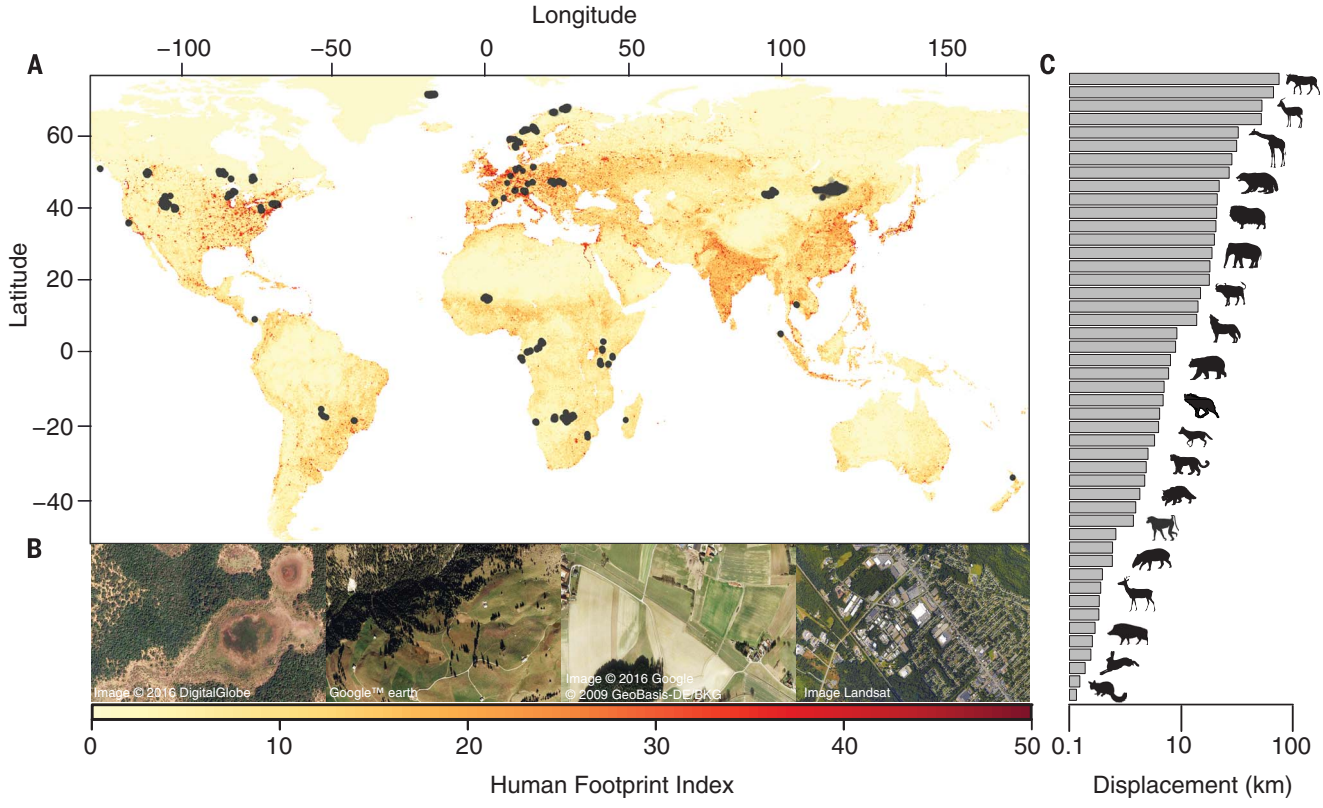


Fig. 1. Locations from the GPS tracking database and the Human Footprint Index. (**A**) GPS relocations of 803 individuals across 57 species plotted on the global map of the Human Footprint Index (HFI) spanning from 0 (low; yellow) to 50 (high; red). (**B**) Examples of landscapes under HFI = 2 (the Pantanal, Brazil), HFI = 20 (Bernese Alps, Switzerland), HFI = 30 (Freising, Germany), and HFI = 42 (Albany, New York). (**C**) Species averages of 10-day long-distance displacement (0.95 quantile of individual displacements). Species (from top to bottom): Mongolian wild ass (*Equus hemionus hemionus*), Mongolian gazelle (*Procapra gutturosa*), giraffe (*Giraffa camelopardalis*), wolverine (*Gulo gulo*), muskox (*Ovibos moschatus*), African forest elephant (*Loxodonta africana cyclotis*), African buffalo (*Syncerus caffer*), wolf (*Canis lupus*), brown bear (*Ursus arctos*), maned wolf (*Chrysocyon brachyurus*), coyote (*Canis latrans*), leopard (*Panthera pardus*), wildcat (*Felis silvestris*), yellow baboon (*Papio cynocephalus*), tapir (*Tapirus terrestris*), roe deer (*Capreolus capreolus*), wild boar (*Sus scrofa*), European hare (*Lepus europaeus*), brushtail possum (*Trichosurus vulpecula*).

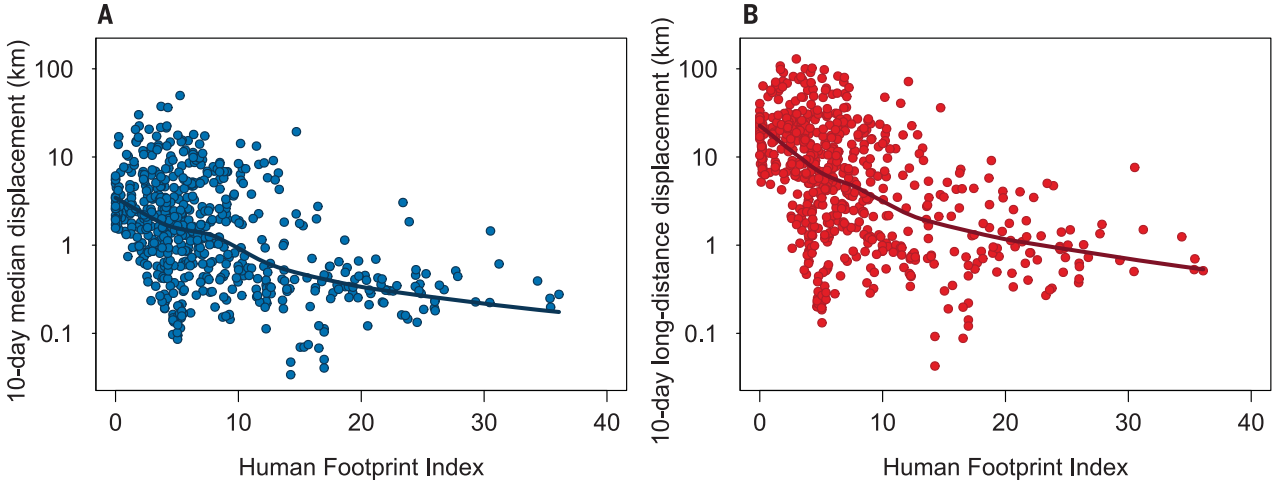


Fig. 2. Mammalian displacement in relation to the Human Footprint Index. (**A**) Median displacements; (**B**) long-distance (0.95 quantile) displacements. Both displacements decline with increasing HFI at the 10-day scale (n = 48 species and 624 individuals). Plots include a smoothing line from a locally weighted polynomial regression. An HFI value of 0 indicates areas of low human footprint; a value of 40 represents areas of high human footprint.

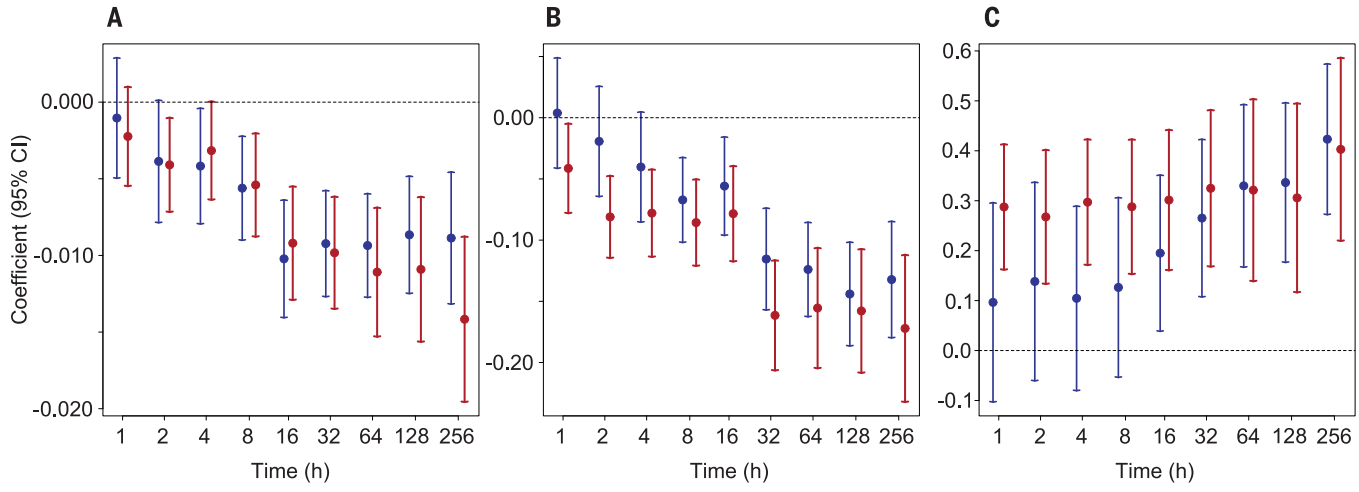


Fig. 3. Model coefficients (with confidence intervals) of linear mixed-effects models predicting mammalian displacements. Coefficient values are shown for (**A**) Human Footprint Index (HFI), (**B**) Normalized Difference Vegetation Index (NDVI), and (**C**) body mass. Models were run for the median (blue) and long-distance (0.95 quantile; red) displacements of each individual calculated across different time scales. Where the error bars cross the horizontal line, the effect is not significant. See table S3 for details.

relationship between resource availability and displacement distance, such that movements were on average shorter in environments with higher resources (Fig. 3B and tables S3 and S4). These results are consistent with reports of larger home range size (*17*) and longer migration distance (*18*) in mammals living in resource-poor environments. Finally, our analyses showed that carnivores traveled on average farther per unit time than herbivores and omnivores (tables S3 and S4). These results concur with prior understanding that carnivores have larger home range sizes (*14*) because they need to find mobile prey and compensate for energy conversion loss through the food web. For all of these variables, effects were significant across time scales longer than 8 hours for both median and long-distance displacements.

The reduction of mammalian movements in areas of high HFI likely stems from two nonexclusive mechanisms: (i) movement barriers such as habitat change and fragmentation (*19*, *20*) and (ii) reduced movement requirements attributable to enhanced resources [e.g., crops, supplemental feeding, and water sources (*5*, *21*)]. Studies have shown both mechanisms at work with varying responses across populations or species (see table S5 for examples). In some cases, they act together on single individuals or populations. For example, red deer in Slovenia have smaller home ranges because of the enhancement of resources via supplemental feeding and the disturbance and fragmentation caused by the presence of roads (*22*).

Although these mechanisms can have differential effects on population densities [i.e., increases under supplementation (*23*) and decreases under fragmentation (*24*)], the consequences of reduced vagility affect ecosystems regardless of the underlying mechanisms and go far beyond the focal individuals themselves. Animal movements are essential for ecosystem functioning because they act as mobile links (*25*) and mediate key processes such as seed dispersal, food web dynamics (including herbivory and predator-prey interactions), and metapopulation and disease dynamics (*26*). Single-species or single-site studies have shown the severe effects of reduced vagility on these processes (*27*, *28*). The global nature of reduced vagility across mammalian species that we demonstrate here suggests consequences for ecosystem functioning worldwide. Future landscape management should strive to maintain landscape permeability by including animal movement as a key conservation metric. Ultimately, because of the critical role of animal movement in human/wildlife coexistence (*29*) and disease spread (*30*), the effects of reduced vagility may go beyond ecosystem functioning to directly affect human well-being.

REFERENCES AND NOTES

1. A. D. Barnosky *et al.*, *Nature* **486**, 52–58 (2012).
2. J. A. Foley *et al.*, *Science* **309**, 570–574 (2005).
3. H. Sawyer *et al.*, *J. Appl. Ecol.* **50**, 68–78 (2013).
4. S. Saïd, S. Servanty, *Landsc. Ecol.* **20**, 1003–1012 (2005).
5. S. Prange, S. D. Gehrt, E. P. Wiggers, *J. Mammal.* **85**, 483–490 (2004).
6. B. Jedrzejewska, H. Okarma, W. Jedrzejewski, L. Milkowski, *J. Appl. Ecol.* **31**, 664–676 (1994).
7. L. A. Tigas, D. H. Van Vuren, R. M. Sauvajot, *Biol. Conserv.* **108**, 299–306 (2002).
8. J. Lenz *et al.*, *Proc. R. Soc. London Ser. B* **278**, 2257–2264 (2011).
9. O. Venter *et al.*, *Nat. Commun.* **7**, 12558 (2016).
10. T. Mueller *et al.*, *Glob. Ecol. Biogeogr.* **20**, 683–694 (2011).
11. N. Pettorelli *et al.*, *Clim. Res.* **46**, 15–27 (2011).
12. W. Jetz, C. Carbone, J. Fulford, J. H. Brown, *Science* **306**, 266–268 (2004).
13. B. K. McNab, *Ecol. Monogr.* **56**, 1–19 (1986).
14. M. A. Tucker, T. J. Ord, T. L. Rogers, *Glob. Ecol. Biogeogr.* **23**, 1105–1114 (2014).
15. J. M. Rowcliffe, C. Carbone, R. Kays, B. Kranstauber, P. A. Jansen, *Methods Ecol. Evol.* **3**, 653–662 (2012).
16. A. M. Hein, C. Hou, J. F. Gillooly, *Ecol. Lett.* **15**, 104–110 (2012).
17. N. Morellet *et al.*, *J. Anim. Ecol.* **82**, 1326–1339 (2013).
18. C. S. Teitelbaum *et al.*, *Ecol. Lett.* **18**, 545–552 (2015).
19. J. F. Kamler *et al.*, *J. Mammal.* **84**, 989–995 (2003).
20. L. Fahrig, *Funct. Ecol.* **21**, 1003–1015 (2007).
21. J. D. Jones *et al.*, *Ecol. Appl.* **24**, 1769–1779 (2014).
22. K. Jerina, *J. Mammal.* **93**, 1139–1148 (2012).
23. J. S. Gilchrist, E. Otali, *Can. J. Zool.* **80**, 1795–1802 (2002).
24. A. Benítez-López, R. Alkemade, P. A. Verweij, *Biol. Conserv.* **143**, 1307–1316 (2010).
25. J. Lundberg, F. Moberg, *Ecosystems* **6**, 87–98 (2003).
26. S. Bauer, B. J. Hoye, *Science* **344**, 1242552 (2014).
27. I. Hanski, O. Ovaskainen, *Nature* **404**, 755–758 (2000).
28. B. F. Allan, F. Keesing, R. S. Ostfeld, *Conserv. Biol.* **17**, 267–272 (2003).
29. M. D. Graham, I. Douglas-Hamilton, W. M. Adams, P. C. Lee, *Anim. Conserv.* **12**, 445–455 (2009).
30. J. M. Hassell, M. Begon, M. J. Ward, E. M. Fèvre, *Trends Ecol. Evol.* **32**, 55–67 (2017).

ACKNOWLEDGMENTS

Supported by the Robert Bosch Foundation and additional funding sources (see supplementary text). The data reported in this paper are available at datadryad.org (doi: 10.5061/dryad.st350). M.A.T., T.M., K.B.-G., W.F.F., J.M.F., and B.V.M. conceived the manuscript; M.A.T. and T.M. conducted the analyses and wrote the first manuscript draft. Co-authors contributed data sets and assisted with writing the final version of the manuscript.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6374/466/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 and S2
Tables S1 to S5
References (*31*–*89*)

15 February 2017; accepted 11 December 2017
10.1126/science.aam9712

MEMBRANE TARGETING

The ER membrane protein complex is a transmembrane domain insertase

Alina Guna,[1] Norbert Volkmar,[2]* John C. Christianson,[2] Ramanujan S. Hegde[1]†

Insertion of proteins into membranes is an essential cellular process. The extensive biophysical and topological diversity of membrane proteins necessitates multiple insertion pathways that remain incompletely defined. Here we found that known membrane insertion pathways fail to effectively engage tail-anchored membrane proteins with moderately hydrophobic transmembrane domains. These proteins are instead shielded in the cytosol by calmodulin. Dynamic release from calmodulin allowed sampling of the endoplasmic reticulum (ER), where the conserved ER membrane protein complex (EMC) was shown to be essential for efficient insertion in vitro and in cells. Purified EMC in synthetic liposomes catalyzed the insertion of its substrates in a reconstituted system. Thus, EMC is a transmembrane domain insertase, a function that may explain its widely pleiotropic membrane-associated phenotypes across organisms.

The mammalian genome encodes hundreds of tail-anchored (TA) membrane proteins with essential roles in diverse processes such as vesicular trafficking, apoptosis, signal transduction, and lipid biosynthesis (*1*). A single transmembrane domain (TMD) close to the C terminus mediates posttranslational TA protein targeting and membrane insertion. Many TA proteins destined for the endoplasmic reticulum (ER) utilize the conserved TMD recognition complex (TRC) targeting pathway whose central component is TRC40 (*2*). Structural studies of Get3, the yeast homolog of TRC40, have revealed a deep hydrophobic groove that binds and shields the hydrophobic TMD of TA proteins (*3*) until their release at an ER-resident receptor complex (*4*, *5*). The surface properties of the substrate-binding groove in Get3 is consistent with biochemical studies showing a preference for TMDs of high hydrophobicity (*6*, *7*). Yet, the TMDs of ER-targeted TA proteins display a wide range of hydrophobicity and length (*1*). Whether or how the TRC pathway might handle this diversity is unclear.

The TMDs from eight ER-destined TA proteins of widely varying biophysical properties (fig. S1) were cloned into a standardized TA protein cassette (Fig. 1A) and shown to insert into ER-derived microsomes in vitro (Fig. 1B and fig. S2). However, only the three most hydrophobic TMDs interacted efficiently with TRC40 by native coimmunoprecipitation (Fig. 1B). Competitive inhibitors of the TRC pathway reduced insertion of only the TA proteins that efficiently engaged TRC40 (Fig. 1C and fig. S3, A to C). The other TA proteins were completely resistant to inhibition. One of these resistant TMDs, from the ER-resident enzyme squalene synthase (SQS), became sensitive to TRC pathway inhibition when the hydrophobicity of its TMD was increased (Fig. 1D and fig. S1). This switch from resistance to sensitivity correlated with TRC40 interaction (Fig. 1E). Even when SQS was assembled with TRC40 in a purified system, the complex dissociated before appreciable insertion into ER microsomes occurred (fig. S4).

These observations indicated that the TRC pathway only handles relatively hydrophobic ER-destined TA proteins. Based on the approximate threshold for TRC40 dependence, we

[1]Medical Research Council (MRC) Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH, UK. [2]Ludwig Institute for Cancer Research, University of Oxford, Old Road Campus Research Building, Headington, Oxford OX3 7DQ, UK.
*Present address: Cambridge Institute for Medical Research, Hills Road, Cambridge Biomedical Campus, Cambridge CB2 0XY, UK.
†Corresponding author. Email: rhegde@mrc-lmb.cam.ac.uk

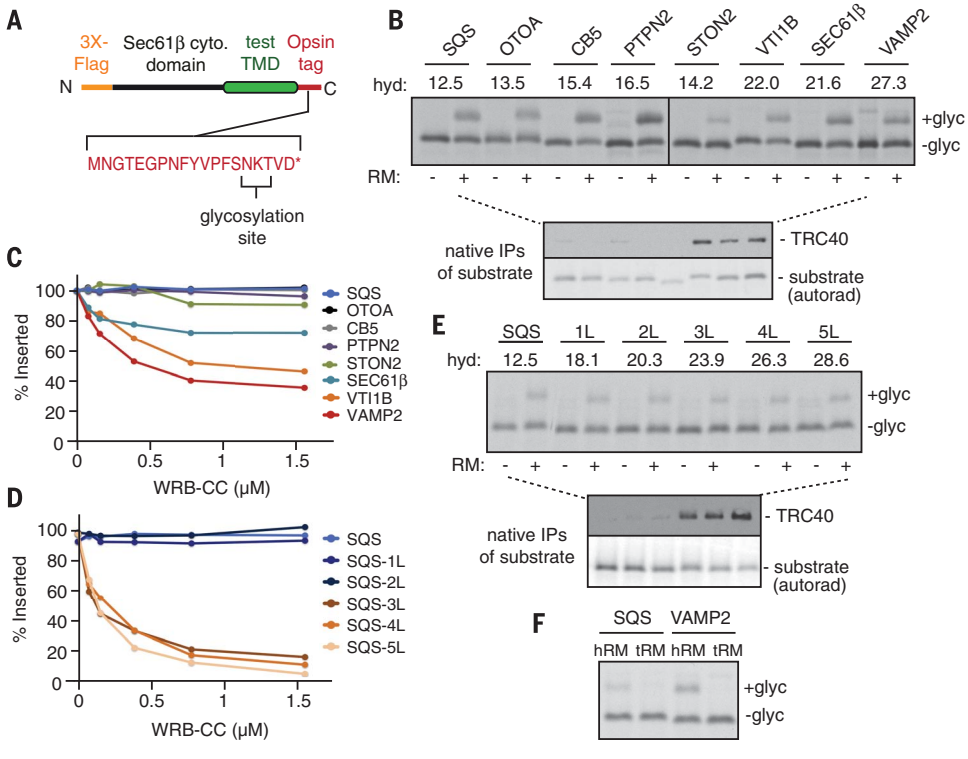


Fig. 1. Detection of a non-TRC insertion pathway for TA proteins. (**A**) Diagram of the TA protein reporter cassette used for most of the analyses in this study. The asterisk at the end of the amino acid sequence indicates the stop codon. (**B**) ^{35}S-methionine–labeled TA protein reporters with the indicated TMDs (see fig. S1) were translated in nucleased reticulocyte lysate (RRL) and incubated with or without canine pancreas–derived rough microsomes (RMs). Glycosylation (+ glyc) indicates successful insertion (see fig. S2). Relative hydrophobicity (hyd) values for each TMD are shown. In a parallel experiment, reactions lacking microsomes for each protein were immunoprecipitated (IP) by means of the substrate's FLAG tag and analyzed for TRC40 association (by immunoblot) and substrate (by autoradiography, autorad). Identical results were obtained in native RRL. (**C** and **D**) Relative normalized insertion efficiencies for the indicated TA proteins with increasing amounts of the coiled-coil domain of the protein WRB (WRB-CC), a fragment of the TRC40 receptor at the ER (see fig. S3A). (**E**) An experiment as in (B) for a set of SQS mutants that successively increase TMD hydrophobicity through leucine (L) residue substitutions (fig. S1). (**F**) Analysis of SQS and VAMP2 insertion using ER microsomes from HEK293 cells (hRM) or trypsin-digested hRM (tRM; see fig. S3D). Single-letter abbreviations for the amino acid residues are as follows: D, Asp; E, Glu; F, Phe; G, Gly; K, Lys; M, Met; N, Asn; P, Pro; S, Ser; T, Thr; V, Val; and Y, Tyr.

estimate that around half of TA proteins are inserted into the ER via a non-TRC pathway. This conclusion is consistent with variable degrees of insertion defects seen when the TRC pathway is impaired (*8*). The mechanism of non-TRC pathway insertion remains unclear, although earlier proposals include unassisted insertion and insertion mediated by the Sec61 translocation channel (*9*, *10*). In support of a protein-mediated process, SQS insertion into ER microsomes pretreated with trypsin was impaired (Fig. 1F and fig. S3D). We thus used SQS as a model non-TRC substrate to identify cytosolic factor(s) that maintain its insertion competence and ER factor(s) needed for its insertion.

Size fractionation and chemical cross-linking were used to compare the cytosolic interactions made by the TMDs of SQS and VAMP2 (vesicle-associated membrane protein 2), an established TRC pathway substrate. As documented previously (*11*, *12*), VAMP2 interacted with each of the factors of the TRC targeting pathway: the chaperone SGTA, the Bag6 quality control complex, and TRC40 (Fig. 2A and fig. S5). The heterogeneous native size of VAMP2, as determined by sucrose gradient fractionation, reflects these multiple interactions (Fig. 2A). By contrast, SQS migrated as a smaller complex and failed to cross-link efficiently to any TRC pathway component (Fig. 2A and fig. S5). The primary cross-link seen with SQS was a ~20 kDa Ca^{2+}-dependent protein (Fig. 2B and fig. S5) that was identified by mass spectrometry as calmodulin (CaM), a factor shown previously to recognize hydrophobic domains in the cytosol (*13*).

Recombinant CaM was sufficient to prevent aggregation of SQS in a chaperone-free *Escherichia coli*–based translation system assembled from purified translation factors (fig. S6). Addition of ER microsomes to the SQS-CaM complex resulted in SQS insertion at efficiencies similar to that observed in total cytosol (Fig. 2C), whereas SQS synthesized in the absence of CaM was aggregated and not insertion competent (fig. S7). SQS insertion occurred concomitantly with release from CaM as monitored by site-specific photo–cross-linking (Fig. 2C). This suggested an insertion model where dynamic substrate release from CaM [at physiologic Ca^{2+} concentrations in the cytosol (*13*)] transiently provides opportunities for ER engagement before recapture by CaM. In support of this model, insertion was precluded if the SQS-CaM complex was stabilized with superphysiologic concentrations of Ca^{2+} (Fig. 2C and fig. S8A), but did occur across the entire physiologic range of cytosolic free Ca^{2+} (fig. S8B). Furthermore, the unrelated TMD chaperone SGTA, which also associates with substrates dynamically (*12*), behaved similarly to CaM in supporting insertion of SQS in both complete cytosol (fig. S9) and purified systems (fig. S10). By contrast, the VAMP2-SGTA complex is insertion incompetent into ER microsomes unless complemented with TRC40 and the Bag6 complex (*12*). Thus, there appears to be a non-TRC pathway tuned to TMDs of moderate to low hydrophobicity. Unlike the highly coordinated TRC targeting system (*2*, *12*), the alternative route can utilize any TMD-shielding factor capable of dynamically releasing substrate for attempts at membrane insertion (Fig. 2D). In native cytosol, the primary factor is CaM (fig. S11), although SGTA can substitute in its absence.

Trypsin sensitivity of the SQS insertion reaction (Fig. 1F) suggested that this critical step is protein mediated. Taking a candidate approach, we considered factors that are conserved across eukaryotes, are abundant, and cause pleiotropic membrane-associated phenotypes when deleted.

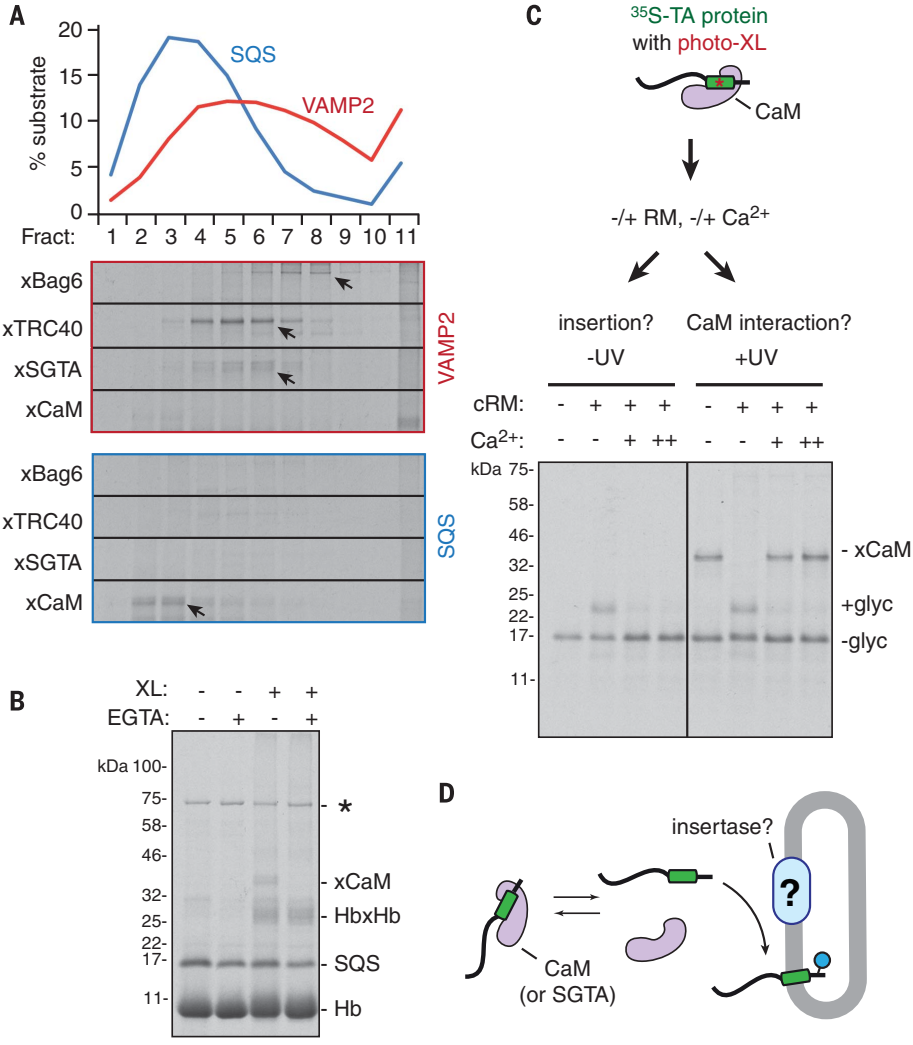
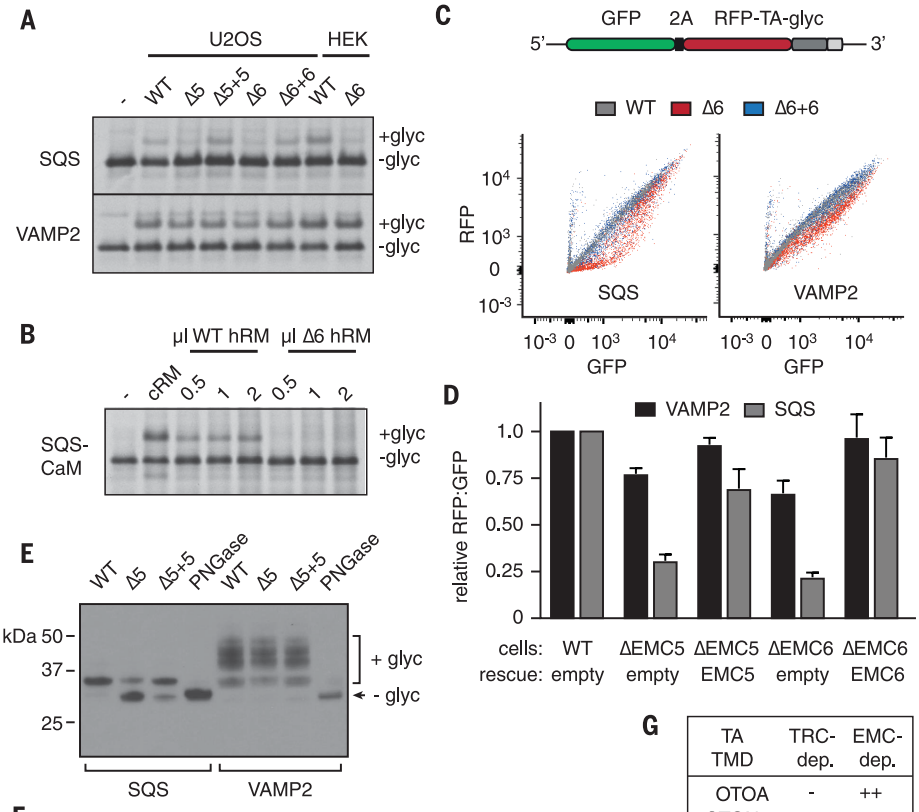


Fig. 2. Identification of cytosolic factors that maintain TA protein insertion competence. (**A**) ^{35}S-methionine–labeled SQS and VAMP2 were translated in native RRL, separated by size on a sucrose gradient, and subjected to chemical cross-linking of each fraction using amine- or sulfhydryl-reactive cross-linker (indicated with an x) (see fig. S5 for full gels). The graph shows the densitometry profiles of each substrate across the gradient, and the individual panels show regions of the cross-linking gels for the indicated interaction partners (verified by immunoprecipitation and mass spectrometry). (**B**) ^{35}S-methionine–labeled SQS translated in native RRL was treated with or without 1 mM EGTA before cross-linking and analysis by SDS–polyacrylamide gel electrophoresis and autoradiography. The major SQS cross-linking partner (xCaM) is not seen with EGTA. Hemoglobin (Hb), its intersubunit cross-link (Hb-Hb), and an unspecified translation product (*) are indicated. XL, cross-linker. (**C**) ^{35}S-methionine–labeled SQS containing the benzoyl-phenylalanine photo–cross-linker within the TMD was produced as a defined complex with CaM by using the PURE system (protein expression using recombinant elements; see fig. S6). The isolated SQS-CaM complex, prepared in 100 nM Ca^{2+}, was incubated with RM in the absence and presence of excess Ca^{2+} (either 0.2 or 0.5 mM) and analyzed directly (left) or irradiated with ultraviolet (UV) light to induce cross-linking before analysis (right). The glycosylated (+ glyc) and CaM–cross-linked (xCaM) products are indicated. (**D**) Schematic of the SQS insertion pathway, with a hypothetical membrane factor indicated with a question mark.



G

TA TMD	TRC-dep.	EMC-dep.
OTOA	-	++
STON2	-	+
CB5	-	++
PTPN2	-	++
SEC61B	+	+
VTI1B	++	-
VAMP2	++	-
SQS	-	++
SQS-1L	-	++
SQS-2L	-	++
SQS-3L	++	-
SQS-4L	++	-
SQS-5L	++	-

Fig. 3. The EMC is essential for TA protein insertion in vitro and in cells. (**A**) Semipermeabilized cells (see fig. S13B) from wild-type (WT) and knockout (Δ) cells of the indicated cell lines were tested for insertion of SQS and VAMP2 by using the glycosylation assay. The "–" indicates a control reaction lacking semipermeabilized cells. (**B**) The isolated SQS-CaM complex (fig. S6) was tested for insertion into cRM or different amounts of hRM from WT or ΔEMC6 (Δ6) HEK293 cell lines. (**C**) Flow cytometry analysis of RFP-SQS and RFP-VAMP2, relative to an internal green fluorescent protein (GFP) expression control (see fig. S15A), in WT (gray), ΔEMC6 (red), or ΔEMC6+EMC6 (rescue, blue) cell lines. Although the RFP:GFP ratio remains close to 1 for VAMP2 across a wide range of expression levels in all cell lines, SQS is selectively decreased in ΔEMC6 cells, especially at low expression levels (see fig. S15B for histograms of these data). 2A, viral 2A peptide. (**D**) Tabulated mean RFP:GFP ratios for SQS (gray bars) and VAMP2 (black bars) in the indicated cell lines. The results for each construct were normalized to the value in WT cells and depict mean ± SD from three independent experiments. (**E**) Immunoblots for SQS-RFP and VAMP2-RFP in the indicated cell lines. Loading was normalized to equivalent amounts of GFP expression as determined by flow cytometry. An aliquot of the WT sample digested with peptide *N*-glycosidase (PNGase) is shown as a marker for nonglycosylated substrate. Glycosylation of the ER-resident SQS is limited to the core *N*-glycan, whereas VAMP2 acquires complex glycans because of trafficking through the Golgi. (**F**) Live cell images of GFP-SQS in the indicated cell lines show altered localization in ΔEMC6 cells. In low-expressing cells (yellow arrows), the localization is diffusely cytosolic, whereas punctae, presumably representing aggregates, are seen in high-expressing cells (red arrows). VAMP2 was unchanged in its localization in ΔEMC6 cells (fig. S15C). (**G**) Summary of dependence on either TRC40 (as judged by inhibitory effect of WRB-CC in Fig. 1) or EMC (see fig. S16) for the indicated substrates.

In preliminary experiments, we observed no effect on SQS insertion of Sec61 inhibition or knockdown of Sec62 or Sec63, arguing against these possibilities (fig. S12). Although genes of the SRP-independent (SND) targeting pathway are synthetic lethal with TRC pathway mutants in yeast (*14*), appreciable impairment of TA protein insertion was not seen in yeast or mammalian cells lacking SND genes (*14*, *15*). We then considered the ER membrane protein complex (EMC), a widely conserved eight- to ten-subunit complex of unknown function (*16*–*18*) (fig. S13A). The EMC is genetically implicated in many unrelated membrane-associated processes such as quality control, trafficking, protein maturation, and lipid homeostasis (*17*–*22*), but its biochemical activity has been elusive.

Using semipermeabilized cultured cells as the source of ER (fig. S13B), we initially noticed that SQS insertion was partially impaired when the EMC5 subunit of EMC was depleted with siRNAs (fig. S13C). Ablation of EMC5 or EMC6 expression by gene editing of osteosarcoma U2OS cells (fig. S14) reduced insertion of SQS, but not VAMP2 (Fig. 3A). This deficiency was rescued by re-expression of EMC5 and EMC6 in the respective knockout cell lines. EMC-dependence was also observed when using ER microsomes isolated from human embryonic kidney (HEK) 293 cells either containing or lacking EMC6 (Fig. 3A). This phenotype was seen regardless of whether the substrates were prepared in crude cytosol (Fig. 3A) or provided as defined complexes with CaM (Fig. 3B) or SGTA (fig. S10B).

We exploited the fact that noninserted TA proteins are typically degraded (*23*, *24*) to analyze SQS insertion in cells. A red fluorescent protein (RFP)–tagged TA protein construct was varied to contain the TMD of either SQS or VAMP2 and analyzed for expression by flow cytometry, membrane insertion by glycosylation, and cellular location by microscopy. Relative to the nearly unimpaired RFP-VAMP2, RFP-SQS showed reduced expression (Fig. 3, C and D, and fig. S15, A and B), impaired glycosylation (Fig. 3E), and altered localization (Fig. 3F and fig. S15C) selectively in EMC knockout cells. Thus, in vitro and in cells, SQS insertion into the ER is dependent on EMC, the absence of which causes SQS mislocalization, degradation, and aggregation.

Analysis of six other TA proteins and the five SQS TMD mutants showed that each TRC40-independent substrate is strongly EMC dependent (Fig. 3G and fig. S16). Sec61β, a protein of moderate hydrophobicity, showed partial dependence on both EMC and TRC40, identifying the approximate point of overlap between these two pathways. Thus, the TRC- and EMC-dependent pathways are mostly tuned for TMDs of high and low hydrophobicity, respectively, although other features such as TMD length or helicity may also influence pathway choice. The lower hydrophobicity of clients for the EMC pathway presumably

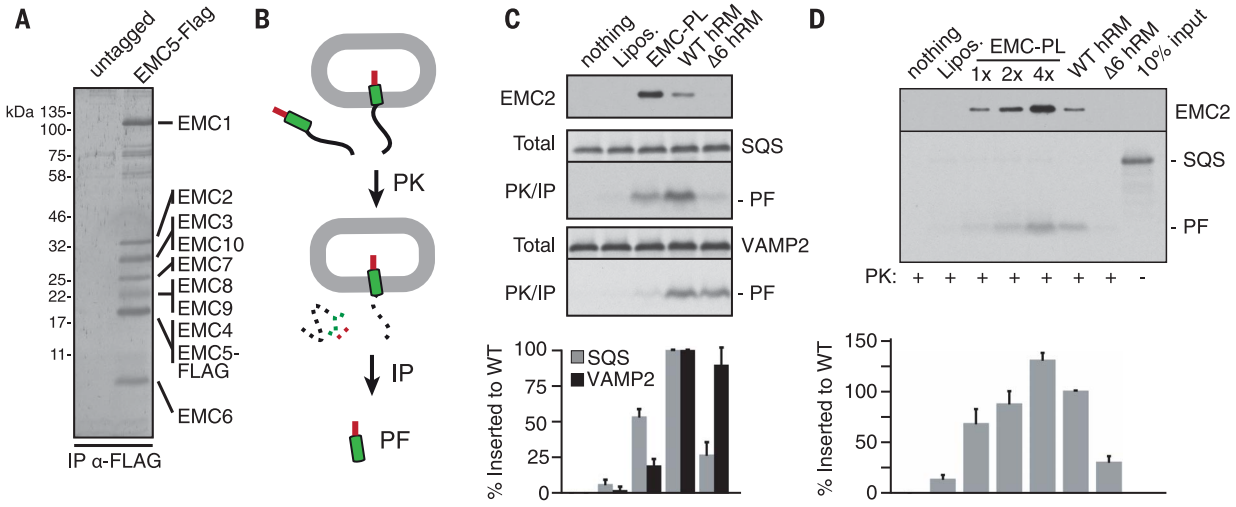


Fig. 4. Reconstitution of EMC-dependent TA protein insertion with purified factors. (**A**) SYPRO Ruby–stained gel of anti-FLAG (α-FLAG) affinity purification from HEK293 cells expressing untagged or FLAG-tagged EMC5. (**B**) Diagram of the protease-protection assay for TA protein insertion using a C-terminal epitope tag (red) to selectively recover the protected fragment (PF) diagnostic of successful insertion. PK, proteinase K; IP, immunoprecipitation. (**C**) Liposomes reconstituted with or without purified EMC were analyzed for insertion of SQS or VAMP2 synthesized in native RRL. For comparison, native ER microsomes (hRM) from WT or ΔEMC6 HEK293 cells were tested in parallel. Immunoblot for EMC2 indicates the relative amounts of EMC. As shown in fig. S18, roughly one-third of EMC in the proteoliposomes is in the correct orientation. The graph represents four experiments (mean ± SD), normalized to insertion in WT hRM. (**D**) Liposomes reconstituted with a constant amount of lipids and varying amounts of purified EMC were analyzed by protease protection for insertion relative to WT and ΔEMC6 hRM. The isolated SQS-CaM complex, an aliquot of which is shown in the last lane, was the substrate for these assays. The samples were also immunoblotted for EMC2 to visualize relative EMC amounts. The graph represents four experiments (mean ± SD) normalized to insertion in WT hRM.

explains why a dedicated targeting pathway with constant TMD shielding is not needed, instead relying on temporary release from general TMD binding proteins to engage the membrane.

To determine whether the EMC is sufficient for TA protein insertion, we purified the intact 10-protein complex (Fig. 4A and fig. S17) and optimized conditions for its reconstitution into liposomes. The reconstituted EMC remained fully intact (fig. S18A), with approximately one-third of the complex oriented correctly (fig. S18B). In a protease-protection assay (Fig. 4B), SQS synthesized in native cytosol inserted into EMC proteoliposomes with approximately half of the efficiency observed in native ER microsomes (Fig. 4C). By contrast, VAMP2 insertion, which is efficient into ER microsomes from both wild-type and EMC6 knockout (ΔEMC6) cells, was poor in EMC proteoliposomes. EMC proteoliposomes also supported insertion of the recombinant SQS-CaM complex at near-native levels of insertion relative to ER microsomes (Fig. 4D) when the amount of correctly oriented EMC was matched (fig. S18, B and C). As expected, SQS insertion was minimal into liposomes (Fig. 4D) or EMC proteoliposomes pretreated with trypsin (fig. S19).

The requirement for EMC in microsomes and in cells for SQS insertion, together with SQS insertion into liposomes at near-native efficiencies by purified EMC, rigorously establishes EMC as an ER-resident insertase for moderately hydrophobic TMDs. Bioinformatic analyses indicate that EMC3 is a distant homolog of Get1 (*25*), a subunit of the insertase for the TRC pathway (*26*). Both Get1 and EMC3 seem to have evolved from an ancestral prokaryotic insertase of the YidC family (*25*), apparently having acquired different substrate specificities in the process. The substrates that fail insertion without EMC probably contribute to many of EMC's reported phenotypes, such as ER stress (*17*), aberrant membrane protein trafficking or degradation (*18*–*21*), altered lipid homeostasis (*22*), or altered viral replication (*27*).

REFERENCES AND NOTES

1. T. Kalbfleisch, A. Cambon, B. W. Wattenberg, *Traffic* **8**, 1687–1694 (2007).
2. R. S. Hegde, R. J. Keenan, *Nat. Rev. Mol. Cell Biol.* **12**, 787–798 (2011).
3. A. Mateja *et al.*, *Science* **347**, 1152–1155 (2015).
4. M. Mariappan *et al.*, *Nature* **477**, 61–66 (2011).
5. S. Stefer *et al.*, *Science* **333**, 758–762 (2011).
6. F. Wang, E. C. Brown, G. Mak, J. Zhuang, V. Denic, *Mol. Cell* **40**, 159–171 (2010).
7. M. Rao *et al.*, *eLife* **5**, e21301 (2016).
8. J. Rivera-Monroy *et al.*, *Sci. Rep.* **6**, 39464 (2016).
9. S. Brambillasca *et al.*, *EMBO J.* **24**, 2533–2542 (2005).
10. B. M. Abell *et al.*, *J. Biol. Chem.* **278**, 5669–5678 (2003).
11. M. Mariappan *et al.*, *Nature* **466**, 1120–1124 (2010).
12. S. Shao, M. C. Rodrigo-Brenni, M. H. Kivlen, R. S. Hegde, *Science* **355**, 298–302 (2017).
13. S. Shao, R. S. Hegde, *Cell* **147**, 1576–1588 (2011).
14. N. Aviram *et al.*, *Nature* **540**, 134–138 (2016).
15. S. Haßdenteufel *et al.*, *FEBS Lett.* **591**, 3211–3224 (2017).
16. J. G. Wideman, *F1000Res.* **4**, 624 (2015).
17. M. C. Jonikas *et al.*, *Science* **323**, 1693–1697 (2009).
18. J. C. Christianson *et al.*, *Nat. Cell Biol.* **14**, 93–105 (2011).
19. M. Richard, T. Boulin, V. J. P. Robert, J. E. Richmond, J.-L. Bessereau, *Proc. Natl. Acad. Sci. U.S.A.* **110**, E1055–E1063 (2013).
20. T. Satoh, A. Ohba, Z. Liu, T. Inagaki, A. K. Satoh, *eLife* **4**, (2015).
21. R. J. Louie *et al.*, *Genome Med.* **4**, 103 (2012).
22. S. Lahiri *et al.*, *PLOS Biol.* **12**, e1001969 (2014).
23. T. Hessa *et al.*, *Nature* **475**, 394–397 (2011).
24. E. Itakura *et al.*, *Mol. Cell* **63**, 21–33 (2016).
25. S. A. Anghel, P. T. McGilvray, R. S. Hegde, R. J. Keenan, *Cell Rep.* **21**, 3708–3716 (2017).
26. F. Wang, C. Chan, N. R. Weir, V. Denic, *Nature* **512**, 441–444 (2014).
27. G. Savidis *et al.*, *Cell Rep.* **16**, 232–246 (2016).

ACKNOWLEDGMENTS

We are grateful to M. Skehel and his team for mass spectrometry analysis; P. Chitwood and B. Phillips for discussions and help with microsome and EMC preparation; R. Voorhees and S. Shao for useful discussions; S. Juszkiewicz for help with flow cytometry analysis; M. Kivlen for help preparing the PURE translation system; and R. Keenan, E. Miller, and S. Juszkiewicz for critical reading of this manuscript. This work was supported by the UK Medical Research Council (grants MC_UP_A022_1007 to R.S.H. and MR/L001209/1 to J.C.C.) and the Ludwig Institute for Cancer Research (J.C.C.). A.G. was supported by a Gates Cambridge Scholarship from the Gates Foundation. Additional data can be found in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6374/470/suppl/DC1
Materials and Methods
Figs. S1 to S19
References (*28*–*44*)

6 July 2017; resubmitted 26 October 2017
Accepted 27 November 2017
Published online 14 December 2017
10.1126/science.aao3099

TECHNICAL COMMENT

ADAPTATION

Comment on "Precipitation drives global variation in natural selection"

Isla H. Myers-Smith[1]* and Judith H. Myers[2]

Siepielski *et al*. (Reports, 3 March 2017, p. 959) claim that "precipitation drives global variation in natural selection." This conclusion is based on a meta-analysis of the relationship between climate variables and natural selection measured in wild populations of invertebrates, plants, and vertebrates. Three aspects of this analysis cause concern: (i) lack of within-year climate variables, (ii) low and variable estimates of covariance relationships across taxa, and (iii) a lack of mechanistic explanations for the patterns observed; association is not causation.

Attribution or testing for mechanistic relationships between climate and ecological or evolutionary parameters is a major challenge in ecology and evolution (*1*–*3*). Siepielski *et al.* (*4*) combined spatial and temporal studies to test whether variation of standardized directional selection gradients and differentials of a variety of traits relate to climate variables at the location of study. Three aspects of their analysis cause us to doubt their overall conclusion:

1) The use of annual gridded climate data by Siepielski *et al*. obscures intra-annual climate variation. We question five aspects of the climate data used in this analysis:

(i) Siepielski *et al.* use annual climate variables that are not relevant to the seasonal periods of the selection experiments or to the temporal scales at which selection acts (*5*, *6*). For example, phenology of egg laying of breeding birds will respond to specific seasonal, not annual, climate variables (*7*). Thus, annual estimates cannot be meaningfully compared across taxa and selection estimates (Figs. 1 and 2).

(ii) Biogeographic patterns could override mechanistic relationships between climate and selection. Both the mean and standard deviation of climate variables used in this analysis are influenced by the geographic locations of the study (Fig. 1, B, D, and F), because the variance of annual temperature is related to latitude and because seasonal variation is greater in colder and higher (or lower) latitudes versus equatorial and warmer locations. Similarly, the variance in precipitation and potential evapotranspiration (PET) is related to site aridity. Greater variation in annual precipitation and PET occurred at sites in temperate regions, particularly for plant taxa (Fig. 1, D and F), which likely influenced the covariance relationships between precipitation variables and selection in plant taxa (Fig. 2, I to L).

[1]School of GeoSciences, University of Edinburgh, Edinburgh EH9 3FF, UK. [2]Biodiversity Research Centre, University of British Columbia, Vancouver, BC V6T 1Z4, Canada.
*Corresponding author. Email: isla.myers-smith@ed.ac.uk

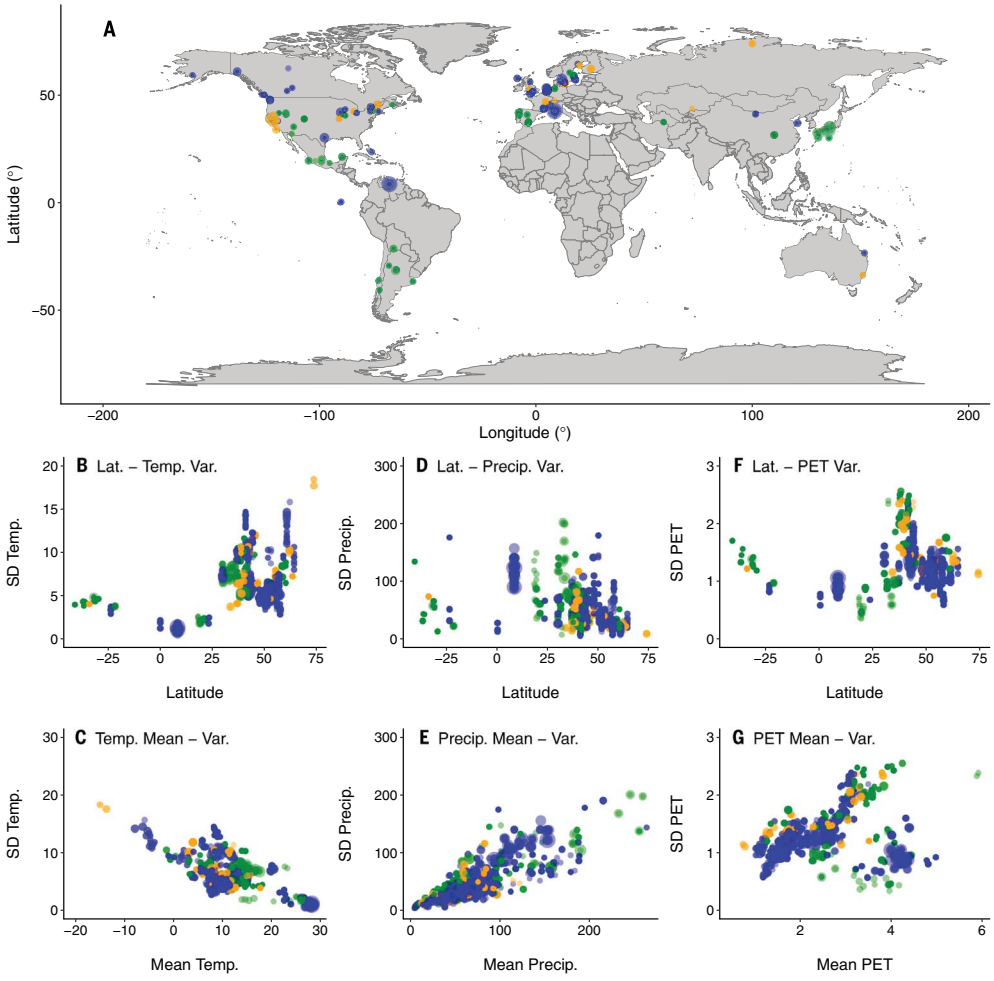


Fig. 1. Map of geographic locations of studies included in Siepielski *et al*. and biogeographic patterns between climate and selection. (**A**) Map of the study locations. (**B** to **G**) Latitude-variance and mean-variance relationships for temperature (°C) [(B) and (C)], precipitation (mm) [(D) and (E)], and PET [(F) and (G)]. A very strong pattern of higher variance in temperature at high latitude and colder locations is clear, and precipitation and PET are variable among geographic locations. Points are colored by taxa (orange, invertebrates; green, plants; blue, vertebrates). The relative size of the points indicates the standard error of the selection estimates.

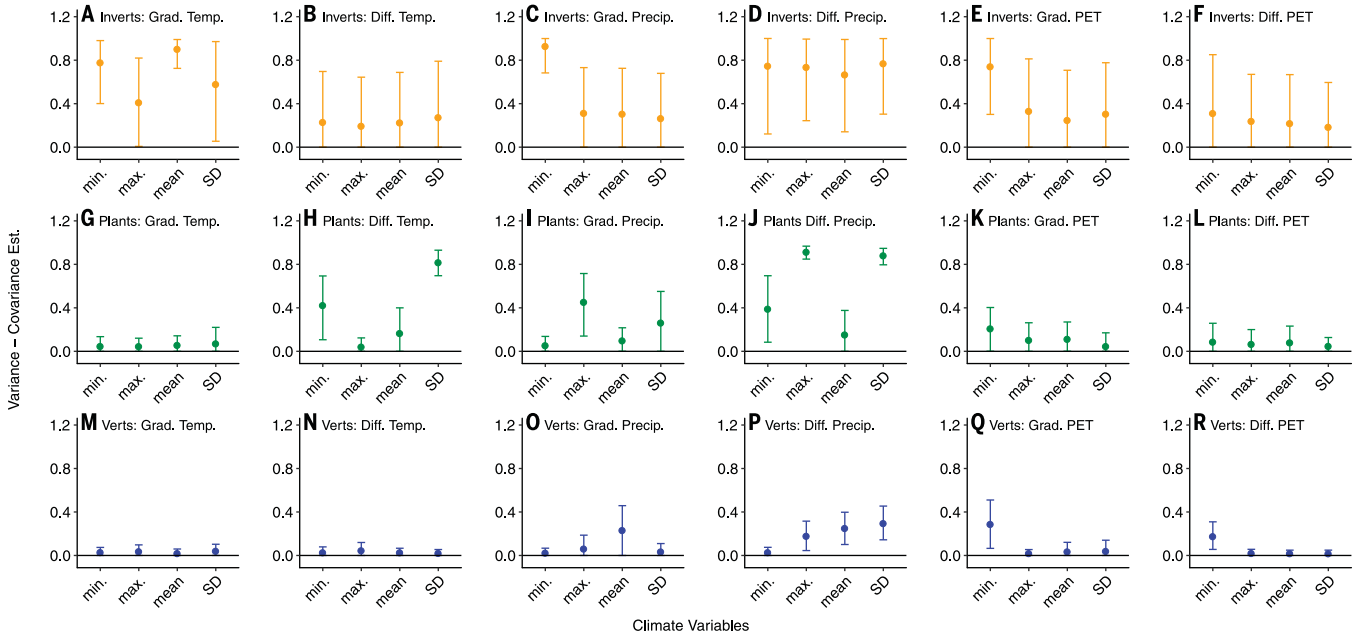


Fig. 2. Taxa-level reanalysis illustrating the high variability in the results among different taxa. (**A** to **R**) Points are the mean proportion of within-study variation in selection; error bars are 95% credible intervals. Climate variables include annual temperature [(A), (B), (G), (H), (M), and (N)], precipitation [(C), (D), (I), (J), (O), and (P)], and potential evapotranspiration [(E), (F), (K), (L), (Q), and (R)]. Color code is the same as in Fig. 1.

(iii) Gridded Climatic Research Unit (CRU) precipitation data are less reliable than temperature data because small-scale variation in landscapes can strongly influence rainfall patterns locally (*8*–*11*). The resolution of climate variables used in this study (0.5° × 0.5°) is much larger than the within-study spatial variation (*4*). This spatial scale mismatch could substantially influence results and is a caveat of many studies working with gridded data at global scales.

(iv) Variance of climate data is lacking at certain sites for certain climate variables; for example, PET data are lacking for 23 of the 165 study sites. When CRU data are not available for a given geographic location for a period of time, a repeated temporal mean value is used in the data set instead (*8*). In particular, PET in the CRU data set is calculated from the gridded mean, minimum and maximum temperature, vapor pressure, and cloud cover; thus, missing data in any of these variables will influence PET data availability (*12*). Care must be taken with gridded data sets to check that data are available and meaningful across all sites and time steps.

(v) The influences of the North Atlantic Oscillation and Oceanic Niño Index, also considered by Siepielski *et al.*, are not consistent in their influences among geographic regions of the world (*13*). Thus, the use of these geographically heterogeneous indices for analyses is not mechanistically interpretable.

2) Siepielski *et al.* find low and variable estimates of covariance relationships across taxa. The taxa-level analyses do not support the overall conclusion of precipitation explaining variation in selection. For invertebrates, temperature measures explain considerable variation in selection, yet with low confidence around those estimates; for plants, relationships of precipitation and within-study variation are inconsistent; and for vertebrates, no associations are apparent (Fig. 2). In addition, within-study variation in selection estimates varied among selection measures. These differences should have been better highlighted in the overall findings of the paper, particularly given that invertebrates make up ~80% of all described species on Earth (*14*).

3) Siepielski *et al.* do not provide adequate mechanistic explanations for the patterns observed. In our assessment, the observed association presented by Siepielski *et al.* should not be interpreted as causation. This study does not present sufficient mechanistic explanations for why annual precipitation or temperature at study locations should be drivers of natural selection across taxa and diverse measures of selection.

For example, Study ID 68 (*15*) tests selection in beak and body size of song sparrows (*Melospiza melodia*) on Mandarte Island, British Columbia, Canada, from 1975 to 1979. The traits included in this study are overwinter survival and reproductive success. Selection in this system is thus likely related to winter mortality (extreme events) or to spring and early summer climate when breeding occurs.

The authors shared some of our concerns about their analysis, such as the temporal mismatch between climate data and selection measures, the low replication, and the low within-study variation (*4*). We believe, however, that these words of caution contradict the confidence with which they present their results.

In summary, for relationships between climate and evolutionary or ecological parameters to be mechanistically meaningful, it is essential that caution be taken with temporal and spatial scales of data, implementation of analyses across diverse taxa, and the interpretation of results. Without appropriate caution, studies have the potential to misdirect future research and understanding.

REFERENCES AND NOTES

1. C. Parmesan *et al.*, *Ecol. Lett.* **16** (suppl. 1), 58–71 (2013).
2. C. Parmesan, *Annu. Rev. Ecol. Evol. Syst.* **37**, 637–669 (2006).
3. A. M. Siepielski, J. D. DiBattista, S. M. Carlson, *Ecol. Lett.* **12**, 1261–1276 (2009).
4. A. M. Siepielski *et al.*, *Science* **355**, 959–962 (2017).
5. S. Kimball, J. R. Gremer, A. L. Angert, T. E. Huxman, D. L. Venable, *Oecologia* **169**, 319–329 (2012).
6. S. J. Franks, S. Sim, A. E. Weis, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1278–1282 (2007).
7. S. J. Thackeray *et al.*, *Nature* **535**, 241–245 (2016).
8. M. Macias-Fauria, A. W. R. Seddon, D. Benz, P. R. Long, K. Willis, *Nat. Clim. Chang.* **4**, 845–846 (2014).
9. F. C. Sperna Weiland, J. A. Vrugt, R. L. P. H. van Beek, A. H. Weerts, M. F. P. Bierkens, *J. Hydrol.* **529**, 1095–1115 (2015).
10. S. Beguería, S. M. Vicente-Serrano, M. Tomás-Burguera, M. Maneta, *Int. J. Climatol.* **36**, 3413–3422 (2016).
11. D. N. Karger, O. Conrad, J. Böhner, T. Kawohl, H. Kreft, R. W. Soria-Auza, N. Zimmermann, H. P. Linder, M. Kessler, Climatologies at high resolution for the Earth's land surface areas, arXiv 1607.00217 (2016).
12. Centre for Environmental Data Analysis, *CRU TS3.21: Climatic Research Unit (CRU) Time-Series (TS) Version 3.21 of High Resolution Gridded Data of Month-by-Month Variation in Climate (Jan. 1901–Dec. 2012)* (2013).
13. N. C. Stenseth *et al.*, *Proc. R. Soc. London Ser. B* **270**, 2087–2096 (2003).
14. B. R. Scheffers, L. N. Joppa, S. L. Pimm, W. F. Laurance, *Trends Ecol. Evol.* **27**, 501–510 (2012).
15. D. Schluter, J. N. M. Smith, *Evolution* **40**, 221–231 (1986).

ACKNOWLEDGMENTS

We thank A. M. Siepielski and M. B. Morrissey for very productive and generous discussions about their work, J. Hadfield for assistance with the statistical analyses, an anonymous reviewer for helpful suggestions, and M. Tseng and the Biodiversity Discussion Group for calling this paper to our attention. All code and data are archived in the following GitHub repository: https://github.com/IslaMS/Siepielski_comment.

22 April 2017; accepted 18 August 2017
10.1126/science.aan5028

TECHNICAL RESPONSE

ADAPTATION

Response to Comment on "Precipitation drives global variation in natural selection"

Adam M. Siepielski,[1]* Michael B. Morrissey,[2] Mathieu Buoro,[3]† Stephanie M. Carlson,[3] Christina M. Caruso,[4] Sonya M. Clegg,[5] Tim Coulson,[6] Joseph DiBattista,[7] Kiyoko M. Gotanda,[6,8] Clinton D. Francis,[9] Joe Hereford,[10] Joel G. Kingsolver,[11] Kate E. Augustine,[11] Loeske E. B. Kruuk,[12] Ryan A. Martin,[13] Ben C. Sheldon,[5] Nina Sletvold,[14] Erik I. Svensson,[15] Michael J. Wade,[16] Andrew D. C. MacColl[17]

The comment by Myers-Smith and Myers focuses on three main points: (i) the lack of a mechanistic explanation for climate-selection relationships, (ii) the appropriateness of the climate data used in our analysis, and (iii) our focus on estimating climate-selection relationships across (rather than within) taxonomic groups. We address these critiques in our response.

Our conclusion that precipitation drives variation in natural selection (*1*) was based on a meta-analysis of selection estimates in wild populations, wherein we estimated how much within-study variation in selection could be explained by climate factors. Myers-Smith and Myers (*2*) have levied three critiques, which we address below.

1) Myers-Smith and Myers point to a lack of mechanistic explanation for climate-selection relationships. This lack of mechanism would be problematic if our goal was to infer causation, or to determine whether selection is more (or less) positive for large or small values of climate variables. However, our goal was to estimate how much within-study variation in selection could be explained by climate. Consequently, we used a random slopes meta-analysis that accommodates a diversity of climate-selection relationships that may arise from many mechanisms. To illustrate, Fig. 1 depicts hypothetical selection coefficients (points) from different studies (colors). In Fig. 1A, there is among-study variation (heterogeneity in intercepts) and within-study variation (residual variation around study-specific regressions), but no climate dependence. In Fig. 1B, there is also within- and among-study variation in selection, but there is a component of the within-study variation in selection (heterogeneity in study-specific slopes) that is associated with climate. The proportion of within-study variance in selection associated with heterogeneity in slopes of climate-selection regressions effectively summarizes climate-selection relationships, irrespective of mechanism. Although our goal was not to determine the extent to which a particular mechanism explains climate-selection relationships, we agree that determining the underlying mechanism(s) is a worthwhile goal.

2) Myers-Smith and Myers suggest that the temporal and spatial scale of the climate data (*3*) was inappropriate because it is not relevant to the temporal and spatial scale at which the studies were conducted. Although higher-resolution data would be preferred, annual estimates can be relevant even if only indirectly linked through correlated climate variables or through correlated environmental responses. More important, Myers-Smith and Myers do not show that the climate measures we used are irrelevant or that they lead to erroneous estimates of climate-selection coupling. As we discussed (*1*), annual measures of climate variables may not support documentation of climate-selection relationships in certain circumstances. Similarly, there are indeed cases where there is no resolvable variation for some variables, particularly potential evapotranspiration (PET). But when this occurs, then such a study does not contribute to meta-analytic inference. This scenario does not detract from the inferences that can be made, because when predictor variables do not reflect causal variables, regression slopes are attenuated (*4*). Consequently, these scale limitations render our analyses conservative but not irrelevant.

Myers-Smith and Myers also note that the climate variables included in our analysis are influenced by geography, and many of these variables are correlated [figure 2 in (*2*)]. This is correct, and we noted the existence of these correlations [table S1 in (*1*)] and their consequences for interpretation. Such correlations among climate variables and their geographic distributions simply reflect the range of climate variables in the studies (*1*). However, because we were interested in within-study climate-selection associations, the relevant associations of climate variables are the within-study variation, not the total variation

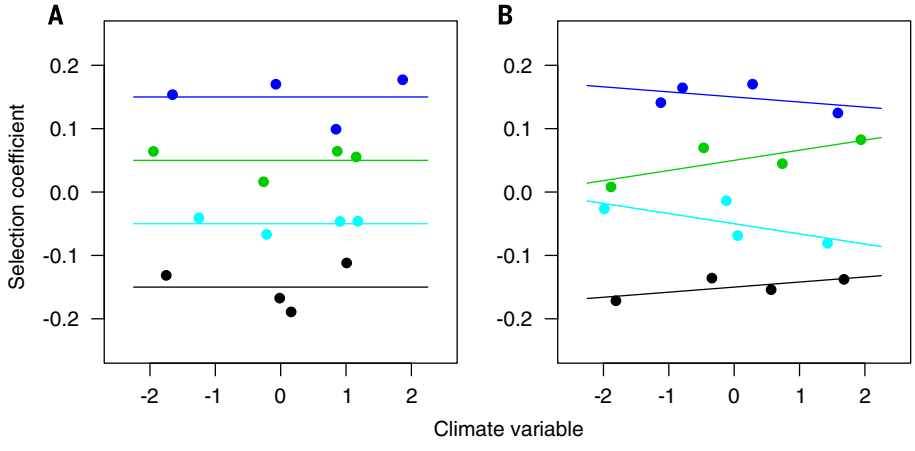


Fig. 1. Two scenarios illustrating how the random slopes mixed model depicts differences in associations between natural selection and climate. (**A**) Within-study (and among-study) variation in selection has no dependence on a climate variable. (**B**) Within-study variation in selection has a component associated with a climate variable.

[1]Department of Biological Sciences, University of Arkansas, Fayetteville, AR, USA. [2]School of Biology, University of St Andrews, St Andrews, UK. [3]Department of Environmental Science, Policy and Management, University of California, Berkeley, CA, USA. [4]Department of Integrative Biology, University of Guelph, Guelph, Ontario, Canada. [5]Edward Grey Institute, Department of Zoology, University of Oxford, Oxford, UK. [6]Department of Zoology, University of Cambridge, Cambridge, UK. [7]Department of Environment and Agriculture, Curtin University, Perth, WA, Australia. [8]Redpath Museum and Department of Biology, McGill University, Montreal, Quebec, Canada. [9]Department of Biological Sciences, California Polytechnic State University, San Luis Obispo, CA, USA. [10]Department of Evolution and Ecology, University of California, Davis, CA, USA. [11]Department of Biology, University of North Carolina, Chapel Hill, NC, USA. [12]Research School of Biology, Australian National University, Canberra, ACT, Australia. [13]Department of Biology, Case Western Reserve University, Cleveland, OH, USA. [14]Department of Ecology and Genetics, Uppsala University, Norbyvägen, Uppsala, Sweden. [15]Department of Biology, Lund University, Lund, Sweden. [16]Department of Biology, Indiana University, Bloomington, IN, USA. [17]School of Life Sciences, University of Nottingham, Nottingham, UK.
***Corresponding author. Email: amsiepie@uark.edu** †Present address: Institut National de la Recherche Agronomique, Université de Pau et des Pays de l'Adour, St Pée sur Nivelle, France.

[i.e., figure 2 in (*2*)]. Within-study climate correlations exist, but they do not preclude our ability to investigate if climate can explain selection.

In addition, Myers-Smith and Myers note that climate indices such as the North Atlantic Oscillation (NAO) are manifested differently in different regions—a point we also made (*1*). If our analysis were based on a global effect (e.g., selection being more positive for high NAO values), this would be problematic. However, within our analytical framework, the action of a climate variable for a given study in one location (i.e., blue line in Fig. 1B) can be completely different for another study (i.e., the green line) in the same, or in a different, location. Both can explain variation in selection even if they act through very different mechanisms.

3) Myers-Smith and Myers argue that our global conclusion about the role of precipitation was problematic because the statistical associations between climate and selection are often weak, and an analysis among taxonomic groups [figure 2 in (*2*); table S5 in (*1*)] shows heterogeneity. We acknowledge that the associations between climate and selection are estimated with much uncertainty (*1*). However, we disagree with the argument that heterogeneity among taxa negates conclusions about effects drawn across taxa. We focused on interpreting the results from the full data set analysis because it is increasingly apparent that zeroing in on a subset of results from many noisy estimates is often the source of the so-called replication crisis (*5*–*7*). In total, there are 126 individual subset analyses of climate-selection coupling estimates [tables S3 to S5 in (*1*)], and among these, thousands of comparisons could be made. We agree that variation among taxa is biologically interesting and that some readers might choose to place more emphasis on the subset analyses.

Myers-Smith and Myers [see also (*8*)] interpret our paper to claim that precipitation, but not temperature, drives variation in selection. However, we are not claiming that temperature has no role; rather, we find a stronger signal of precipitation. Results from subset analyses of space [fig. S1 in (*1*)], but not time, as well as other subsets [tables S3 to S5 in (*1*)] do show a signal of temperature on selection. We focused our conclusions on the role of precipitation because that conclusion is based on the global analysis. We agree that temperature may very well have important effects, and we acknowledged this (*1*).

In conclusion, we agree with many of the points raised by Myers-Smith and Myers, and indeed noted most of these caveats ourselves (*1*). However, their reanalysis essentially replicates the results of our work and therefore does not conflict with our findings. Instead, their interpretation of the results differs from our main conclusions, in part because of an emphasis on different aspects of the available data. As such, we believe that our conclusion stands that precipitation is a driver of selection in wild populations.

REFERENCES AND NOTES

1. A. M. Siepielski *et al.*, *Science* **355**, 959–962 (2017).
2. I. H. Myers-Smith, J. H. Myers, *Science* **359**, eaan5028 (2018).
3. Centre for Environmental Data Analysis, *CRU TS3.21: Climatic Research Unit (CRU) Time-Series (TS) Version 3.21 of High Resolution Gridded Data of Month-by-Month Variation in Climate (Jan. 1901–Dec. 2012)* (2013).
4. D. T. Campbell, D. A. Kenny, *A Primer on Regression Artefacts* (Guilford, 1999).
5. Open Science Collaboration, *Science* **349**, aac4716 (2015).
6. T. H. Parker *et al.*, *Trends Ecol. Evol.* **31**, 711–719 (2016).
7. N. P. Lemoine *et al.*, *Ecology* **97**, 2554–2561 (2016).
8. M. Tseng, H. Eyster, Both temperature and precipitation drive global variation in natural selection [*Science* eLetter]; science.sciencemag.org/content/355/6328/959/tab-e-letters.

ACKNOWLEDGMENTS

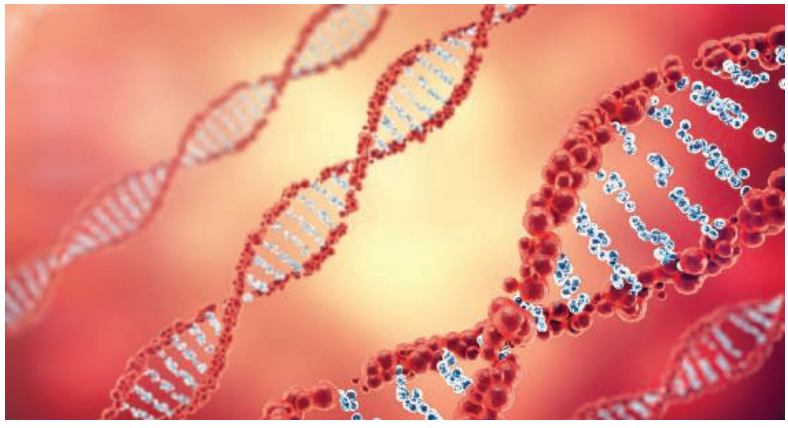
We thank I. H. Myers-Smith and J. H. Myers for discussions of the nature of climate data in ecological and evolutionary studies, and an anonymous reviewer for helpful suggestions. Supported by NSF grant DEB1620046 (A.M.S.).

16 May 2017; accepted 28 August 2017
10.1126/science.aan5760


RESEARCH ON THE EDGE
28.57050°N | 80.6490°W
KENNEDY SPACE CENTER, FL
ONE DAY WE'RE LAUNCHING SATELLITES. THE NEXT WE'RE LAUNCHING CAREERS.
At the University of New Hampshire, we find answers where you wouldn't expect. Like on more than 20 NASA satellites, which carry research from our Space Science Center as far as the edge of our solar system. And because students are essential research partners on these projects, we've managed to launch some stellar careers along the way.
Welcome to the edge of possible: where intellectual curiosity is rewarded with exciting research opportunities and challenging academic programs. Here, students thrive under the mentorship of Guggenheim and Carnegie fellows and winners of the Pulitzer and other honors. Whether it's blasting research into space or training tomorrow's guardians of intellectual property, we exist to help our students create possibilities for our world.
unh.edu/edge
University of New Hampshire

MICHELSON MEDICAL RESEARCH FOUNDATION

HUMAN VACCINES PROJECT



CALL FOR ENTRIES

NEW! Michelson Prizes for Human Immunology and Vaccine Research

The Michelson Prizes are international awards that support young investigators applying innovative research concepts and disruptive technologies to significantly advance the development of future vaccines and immunotherapies.

Proposals will be reviewed by a distinguished committee of internationally recognized scientists.

Researchers under the age of 35 are invited to apply.

Learn more about the Prizes and how to apply at **humanvaccinesproject.org/michelsonprizes**

PRE-APPLICATION DEADLINE FEBRUARY 12, 2018

Be the first to win this prize and receive:

- Prize money of US $150,000 each will be awarded to two investigators
- Present your research at a conference on the Convergence of Human Immunology and Vaccine Research at the Michelson Center for Convergent Bioscience at the University of Southern California

APPLY TODAY!

Flow Cytometer

An evolution of the CytoFLEX platform, the new CytoFLEX LX Flow Cytometer is available with up to six lasers and 21-color detection. It offers great performance and exquisite sensitivity for small-particle analysis in a benchtop analyzer, an extensive set of repositionable bandpass filters, flexibility to upgrade by adding additional parameters, and intuitive software to facilitate multicolor analysis. The fully activated instrument includes three fluorescent channels from the 488-nm (blue) laser, three from the 638-nm (red) laser, five from the 405-nm (violet) laser, five from the 561-nm (yellow) laser, three from the 375-nm (near-UV) laser, and two from the 808-nm (infrared) laser. Go ahead—push the boundaries of flow cytometry with the CytoFLEX LX.

Beckman Coulter Life Sciences
For info: 800-742-2345
www.beckman.com

Autophagy Reagents

Sino Biological offers autophagy-related (Atg) molecules and proteins. Autophagy was the subject of the 2016 Nobel Prize in Physiology or Medicine—it controls important physiological functions through an internal recycling program, eliminating invading intracellular bacteria and viruses, contributing to embryo development and cell differentiation, and functioning in signaling pathways and therapy targets in cancer. It plays a role in innate and adaptive immunity, programmed cell death, neurodegeneration, and aging. Examples of autophagy reagents offered by Sino Biological include Atg molecules and MARCH2, a newly discovered negative regulator of cell autophagy. MARCH2 regulates autophagy by promoting CFTR ubiquitination and degradation, and PIK3CA-AKT-mTOR signaling.

Sino Biological
For info: +86-10-5862-8288
www.sinobiological.com

Secondary Antibodies

Invitrogen Alexa Fluor Plus secondary antibodies represent an advancement in fluorescent-conjugate technology, designed to provide brighter signals, enhanced sensitivity, and minimal cross-reactivity in a variety of applications. Take your cell imaging and Western blotting results to another level—these antibodies offer up to 4x higher signal-to-noise ratio than others on the market; superior detection and visualization of low-abundance targets in challenging samples; improved sensitivity and range of linear detection for more detail; and a unique, proprietary dye chemistry you can't get from anyone else. Make your low-abundance targets visible, spend less time optimizing, and make every one of your precious samples count.

Thermo Fisher Scientific
For info: 800-955-6288
www.thermofisher.com

Protein Aggregate Separation and Characterization System

Postnova Analytics' AF2000 MALS is a temperature-controlled, asymmetrical flow field-flow fractionation (AF4), multiangle light scattering (MALS) system, developed to provide unmatched separation, characterization, and fractionation of biopharmaceutical proteins. The system replaces the liquid chromatography column with a proprietary separation-channel technology that retains molecules based only on size and by the influence of crossflow as a separation force. Because of its unique design and the absence of any stationary phase, separation is achieved without exerting shear forces and stress on proteins and aggregates. Incorporating MALS and AF4, the AF2000 MALS operates over a very broad molar-mass and particle-size range. The system can achieve separations with higher resolution, reproducibility, and recovery than previously possible.

Postnova Analytics
For info: +44-(0)-1885-475007
www.postnova.com/af2000-overview.html

CRISPR Guide RNA

A shortfall of many CRISPR-Cas9 gene-editing experiments is that the guide RNA, which directs the Cas9 nuclease to cut at a specific genomic DNA target site, may cause a double-strand break without triggering functional knockout of the protein. We use a powerful algorithm in the genome-wide design of our Edit-R Predesigned Synthetic crRNAs and Lentiviral sgRNAs, for editing of human, mouse, and rat genes. Now you only need to do a gene search to get either synthetic CRISPR RNA (crRNA) or lentiviral single-guide RNA (sgRNA) products for functional target-gene knockout. By assessing functional phenotypes for thousands of designs, then validating our design rules in other assay systems, we have established rules for determining target sites that are more likely to result in efficient cleavage and functional knockout with high specificity.

GE Healthcare Dharmacon
For info: 800-235-9880
dharmacon.gelifesciences.com

RNA Methylation Analysis

Expand your understanding of RNA methylation with EpiGentek's complete suite of RNA methylation products. Recent data strongly suggest that m6A and 5-mC RNA methylation affects regulation of various biological processes such as RNA stability and mRNA translation, and that abnormal RNA methylation contributes to the etiology of human diseases. From directly quantifying the "fifth RNA base," N6-methyladenosine (m6A), to measuring global 5-methylcytosine (5-mC) RNA and more, our kits make it easy for you to determine how RNA methylation influences your research projects. We offer one m6A RNA kit [EpiQuik m6A RNA Methylation Quantification Kit (Colorimetric)] and three 5-mC RNA kits: MethylFlash 5-mC RNA Methylation ELISA Easy Kit (Fluorometric); EpiNext 5-mC RNA Bisulfite-Seq Easy Kit (Illumina); and Methylamp RNA Bisulfite Conversion Kit.

EpiGentek
For info: 877-374-4368
www.epigentek.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.

MERCK'S POSTDOCTORAL RESEARCH FELLOW PROGRAM

For more than a century, a special passion has driven us at Merck.* Our goal is to discover and develop medicines, vaccines and animal health innovations that will help improve the lives of millions. Still, we know that there is much more to be done. And we're doing it, with a long-standing commitment to research and development.

Developing Talent, Harnessing Innovation, Building for the Future

Our Postdoctoral Research Fellow Program builds on our legacy of scientific excellence and innovation.

If you join us, you'll be a part of a team of motivated scientists working to discover and develop medicines and vaccines that help meet the world's unmet medical needs. You will work alongside outstanding researchers and collaborators as part of Merck's industry-leading research and development organization. Our postdocs:

- Work in unique laboratory environments with top capabilities, equipment, expertise and knowledge
- Obtain experience in drug discovery and development
- Generate innovative science resulting in high-quality external publications
- Participate in seminars, lectures and meetings, and have the opportunity to interact with the local scientific community
- Become positioned for choice careers in pharma, biotechnology or academia

As a postdoc, you will receive a competitive salary and access to the full benefits programs offered by Merck.

At Merck, our passion is improving health. This is what keeps us at the forefront of scientific discovery and innovation. We invite you to apply.

To learn more about the Merck Research Laboratories (MRL) Postdoctoral Research Fellow Program, visit www.merck.com/research/fellow/home.html.

To apply for a position, visit www.merck.com/careers.

*Merck & Co., Inc., Kenilworth, NJ, USA, is known as MSD outside the United States and Canada.

MERCK
INVENTING FOR LIFE

Connect with us on Facebook, Twitter and LinkedIn.

merck careers
merckbewell
merckcareers1
merck

Postdoctoral Research Fellow Program

195421-0002

NEW ENGLAND BioLabs® Inc.

be INSPIRED
drive DISCOVERY
stay GENUINE

We've spread our wings.

Monarch® Nucleic Acid Purification Kits Now available for DNA & RNA

Designed with sustainability in mind, Monarch® Nucleic Acid Purification Kits are the perfect complement to many molecular biology workflows. Available for DNA & RNA purification, with buffers and columns available separately, Monarch kits are optimized for excellent performance, convenience and value. Quickly and easily recover highly pure, intact DNA and RNA in minutes. Available kits include:

- Monarch Plasmid Miniprep Kit
- Monarch DNA Gel Extraction Kit
- Monarch PCR & DNA Cleanup Kit (5 µg)
- **NEW** MONARCH TOTAL RNA MINIPREP KIT – optimized for use with a variety of sample types, including cells, tissues, blood, and more

Make the change and migrate to Monarch today.

Interested in trying a sample of our new Monarch Total RNA Miniprep Kit?

Learn more at
NEBMonarch.com

POSITIONS OPEN

MEDICAL SCHOOL
UNIVERSITY OF MICHIGAN

POSTDOC PREVIEW
May 31 – June 1 2018
University of Michigan, Ann Arbor

Highlights: Fully funded travel, lodging and meals; Interviews with prospective mentors; Information about NIH-funded Training Programs; Networking with UM Postdoctoral Association; Tours of campus and town.

Eligibility: Graduate students who expect to complete their Ph.D. by December 2018 and who are U.S. citizens or permanent residents or who are currently enrolled in a U.S. graduate institution.

Applications open February 12 and close March 26, 2018

More information can be found: on our website at **https://medicine.umich.edu/medschool/education/postdoctoral-fellows/prospective-postdocs/postdoc-preview** or by sending e-mail to **Postdocoffice@med.umich.edu**

Postdoctoral Fellow
Tulane University, School of Medicine,
New Orleans, LA

The Postdoctoral Fellow for the Hua Lu laboratory in the Department of Biochemistry and Molecular Biology will be responsible for studying the biochemical and molecular mechanisms underlying the post-translational regulations of p53 and their role in cancer stem cell growth and proliferation.

Qualifications: Among other qualifications, successful candidates should have a Ph.D. or equivalent degree (within 6 months of hiring) and have prior research experience in biochemistry, enzymology, protein purifications, and molecular biology. See below link for additional details.

Application Instructions: Candidates must apply in Interfolio and provide the following materials for submission: Cover letter, Curriculum Vitae and list of references with contact information.

This institution is using Interfolio's Faculty Search to conduct this search. Applicants to this position receive a free Dossier account and can send all application materials, including confidential letters of recommendation, free of charge.

Apply Now at website - apply.interfolio.com/48163

Tulane University is an Equal Employment Opportunity/Affirmative Action Institution committed to excellence through diversity. Tulane University will not discriminate based upon race, ethnicity, color, sex, religion, national origin, age, disability, genetic information, sexual orientation, gender identity or expression, pregnancy, marital status, military or veteran status, or any other status or classification protected by federal, state, or local law. All eligible candidates are encouraged to apply.

The **Arbeitman Laboratory at Florida State University, College of Medicine** studies the molecular-genetic basis of complex reproductive behaviors (http://neuro.fsu.edu/faculty/arbeitman). Postdoctoral Scholar positions are available for NIH-funded ongoing projects to understand behavior, using a range of approaches, including neural circuit based studies and cutting-edge, single-cell genomic-scale approaches. FSU has initiated several career enhancing programs, including a Provost Postdoctoral Fellowship and Teaching/Research Postdoctoral Fellowship (http://opda.fsu.edu). Interested candidates should send a Curriculum Vitae to Michelle Arbeitman for consideration (e-mail: **michelle.arbeitman@med.fsu.edu**).

北京大学
120th ANNIVERSARY OF PEKING UNIVERSITY
1898–2018

The Institute of Social Science Survey at Peking University

The Institute of Social Science Survey (ISSS) was founded by Peking University in 2006 due to the rapid changes in various aspects of Chinese society and a grave lack of high-quality data to track such changes. It aims to collect micro-data that has both academic and policy values.

ISSS maintains two of China's largest non-government surveys: China Family Panel Studies (CFPS) and China Health and Retirement Longitudinal Study (CHARLS). By adopting the most up-to-date computer-assisted survey techniques, the two projects have conducted multiple rounds of survey and collected the most influential micro-data to date on China's livelihood issues. The two studies, reported by *Science* in 2010 and 2013, have attracted over 40,000 data users worldwide and produced nearly 1,300 papers which appeared in leading journals such as *Lancet, Proceedings of the National Academy of Sciences*, and *Annual Review of Sociology*.

CFPS is a national longitudinal general survey project and aims to document changes in Chinese society by collecting data at the individual, family and community levels. It focuses on both the economic and the non-economic wellbeing of Chinese people, including such topics as economic activities, educational attainment, family dynamics, physical and mental health. CFPS launched its national baseline survey in 2010 and collected data of 42,590 individuals from 14,960 households across 25 provinces in China. CFPS has conducted follow-up surveys every two years on all baseline family members and their newborns and adopted children. CFPS is led by Yu Xie, a member of the US National Academy of Sciences and a professor at Princeton University and Peking University, and Xiaobo Zhang, Ping Tu, and Qiang Ren, who are all faculty members at Peking University.

CHARLS is another national survey, targeting households with members aged 45 years or above and aims to provide a high-quality public micro-data for multidisciplinary study of population aging. It conducted its national baseline survey in 2011-2012 on 17,708 individuals in 10,257 households and covered 150 counties/districts, 450 villages/urban communities across the country. Follow-up surveys were successfully conducted in 2013 and 2015, leading to the release of two reports: "Challenges of Population Aging in China" and "China Health and Retirement Report". CHARLS is led by Yaohui Zhao, professor of economics at Peking University, John Strauss from the University of Southern California and Gonghuan Yang from the Chinese Academy of Medical Sciences.

The Director of ISSS is Professor Qiang Li, a political scientist from Peking University. In response to the increasing demand for data collection projects, he restructured ISSS in 2011 to expand the institute's scope by fully utilizing the team's potentials in survey operation and techniques. Thanks to this restructuring, ISSS has been able to undertake other major surveys. In 2013, ISSS conducted the China Mental Health Survey (CMHS) on nearly 30,000 individuals from 1,256 villages across China. More recently, ISSS piloted a survey on small and medium-sized enterprises in 18 counties, ready to launch its national baseline in 2018.

In addition to data collection, ISSS is a pioneer in promoting the long-term preservation, access and reuse of survey data from big and small surveys in China. It has partnered with the University Library to set up the China Survey Data Archive (CSDA), which now hosts 158 datasets and 574 technical documents.

The institute also plays an active role in supporting empirical research that has policy implications. Ongoing research by the institute's affiliates covers a wide range of topics including income inequality, fertility policy, aging, enterprise reform, and the government's role as a service provider. Research findings are published in the institute's newsletter series Data and Public Policy on a regular basis.

ISSS also embraces the challenges from big data, and has set up a big data lab to explore the possibility of linking big data and survey data for interdisciplinary collaboration among departments inside and outside Peking University.

ISSS welcomes job applicants and visiting scholars with expertise in survey methodology, empirical analysis, big data and related fields. Feel free to contact us:

Website:
www.isss.pku.edu.cn
Email: isss@pku.edu.cn
Tel: 010-62767908
Fax: 010-62759641

The Laboratory of Archaeolgical Science at Peking University

The Laboratory of Archaeolgical Science at Peking University was founded in the 1970s when the radiocarbon dating method was first used at Peking University. Committed to determining the age of archaeological sites, this lab has witnessed the dating of dozens of archaeological finds, particularly those pertaining from the Upper Palaeolithic Period to the Neolithic Period. It was in this lab that the skeletons of Tianyuan Man, Laishui Man and Peking Man were successfully dated, laying the foundation for the research on Homo sapiens in China. Its most recent finding, published in *Science*, is its dating of the earliest pottery (not only in China but also in the world) at 20 ka, subverting the then dominant notion within academia that pottery was first made at 10ka.

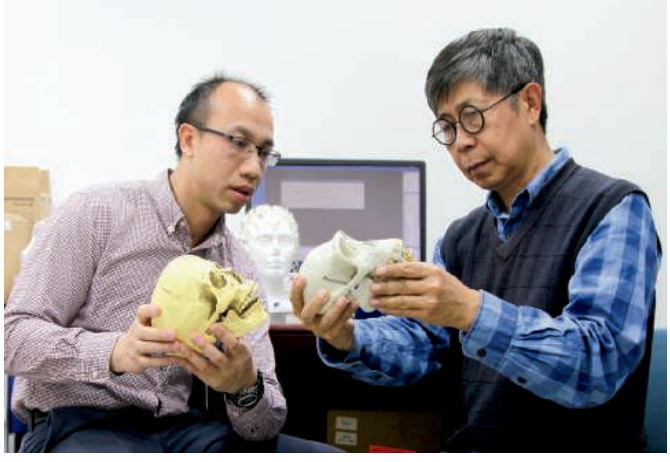
Since 2002, the lab has undertaken several key projects of national importance, pertaining to the origin of Chinese civilization and the early phases of its development in 3,500 BC – 1, 500 BC in the core cultural regions, such as the Central Plain, the middle and lower reaches of Yellow River, the middle and lower reaches of the Yangtze River, and the West Liaohe River Basin. Through 15 years of successive study, about 2,000 Hi-precision radiocarbon dating results have been accomplished so that a more precise chronological framework has been made possible, laying a solid foundation for the research on the evolution of early Chinese civilization and the interaction of its ancient cultural regions.

The lab also carries out archaeo-metallurgical and archaeo-material studies on such artifacts as ceramics, jade and glass to deduce their provenance and manufacturing techniques, providing evidence for the research on the origin of Chinese civilization and the cultural exchange between the East and the West.

We hope for communication and cooperation with colleagues at home and abroad. Please feel free to contact us.

Website:
http://archaeology.pku.edu.cn/
Email: wuxh@pku.edu.cn

The Linguistic Laboratory at Peking University



The Linguistic Laboratory at Peking University, formerly known as the Laboratory of Phonetics and Musics, was established in 1925 by Professor Fu Liu. It is the first interdisciplinary lab of liberal arts and sciences in China.

During the 30s and 40s of last century, the lab focused on the study of tones using mechanical acoustic devices, such as the kymograph; then, in the second half of last century, electronic acoustic devices, such as the spectrograph, were used for speech-sound analysis. Since 2000, instruments including MRI, EEG, and Eye tracker are applied to study speech and oral cultures from not only physiological but also psychological perspectives.

At present, the lab's research covers 4 sub-fields, namely: 1) the speech sounds of languages in China; 2) the physiological speech models of Mandarin; 3) traditional Chinese oral cultures and musicology; 4) the dyslexia of pupils in Beijing.

In recent years, Professor Jiangping Kong's team has been devoted to the research on the physiological models of human vocal tracts based on MRI materials. Through international cooperation, the models in turn form the basis for the studies of the evolution of speech organs from chimpanzees to modern human.

As is well known, vowels [a,i,u] are regarded as the basic vowels in human speech. The chimpanzee, on the other hand, can only produce sounds which resemble [a] and [u], because their larynx has not descended as it has in humans. The descended larynx in humans results in the fronting and raising of the tongue by contraction of the genioglossus muscles and jointly facilitates positioning the tongue for the production of high front vowels. The differentiation of chimpanzee and modern human dates back to 5 million years ago. Since tongue muscles and larynx do not fossilize, it is difficult to reconstruct vocal tracts simply based on the fossils of ancient Homo.

Using 3D MRI data, Kong's team has established the vocal tract models of chimpanzee and modern human, which includes the parameters of front cavity length, back cavity length and shape similarity. Through the vocal tract simulation from that of chimpanzee to that of modern human, about fifty thousand 3D vocal tracts have been reconstructed, together with the resultant synthesized sound samples. By undertaking speech perception tests, the team explores the time vowels could have emerged. Further research on this will deepen our understanding of the evolution of speech organs, as well as the origin of human language.

The lab welcomes international cooperation. Feel free to contact us:

Website:
http://chinese.pku.edu.cn/teachingstaff/linguisticlaboratory/
Email: jpkong@pku.edu.cn

Do I make myself clear?
Media training for scientists

Media savviness previously consisted of on-the-job training. Now, more universities and organizations are offering programs to help scientists get their message across to the public.
By Charlotte Schubert

Ryan Kelly's office reflects his eclectic interests. A poster of chitons, the marine invertebrates he once studied, leans against the wall. On his shelves, books on marine ecology are intermingled with law, policy, and science communications texts.


Ryan Kelly

Kelly, an assistant professor at the University of Washington's (UW) School of Marine and Environmental Affairs, pulls down one book, *Escape from the Ivory Tower: A Guide to Making Your Science Matter*, by **Nancy Baron,** the director of science outreach at COMPASS, a public engagement organization for scientists.

Kelly thumbs through the book, pointing to a one-page chart designed by COMPASS to help scientists organize their thoughts before presenting their research to a broader audience. "Here it is," he says. "The message box."

He explains how he used the message box before media interviews about a recent study. When he first released the paper, he contacted a press officer at UW who told him that the study was likely to be of wide interest. Kelly then wrote his own press release, and after the press officer edited it, began working with the message box.

In the center of the box is a space for the core issue, in Kelly's case, "If you want your science to matter, tell a story." Flanking the center message are spaces to write the problem, its solution, the benefit of the study, and why people should care about it. That exercise helped Kelly prepare to talk with journalists from the BBC and other outlets. "The more prepared you are, the more relaxed you can be. The message box is one way to do that," says Kelly.

When his research on marine ecosystems or marine law puts him in front of policymakers, journalists, or the public, Kelly also leans on training he received through a communications workshop at UW.

Kelly has tapped into a trend. Scientists like him are increasingly gaining communications savvy from formal training, whether through their workplaces or through roving workshops sponsored by organizations such as COMPASS in the United States, or the Science Media Centre in the United Kingdom.

In the last decade or so, an increasing number of scientists, many of them just beginning their careers, have begun to view science communications as integral to their work. "There has been a huge shift," says Baron, who is based in Santa Barbara, California. "It is not 'should I do it?' but '*how* should I do it?'"

Though meaningful numbers are hard to come by, those who run science communications workshops observe that demand has only grown in the last decade, and more institutions worldwide are sprouting homegrown programs. Funders, such as the Rita Allen Foundation and the Burroughs Wellcome Fund, are encouraging their growth.

Communications training now goes far beyond providing media tips, to teaching a general skill set that can be deployed to engage a broad audience through the medium of a scientist's choice, be it public talks, video, social media, or other self-made content.

"We have a lot of individual experiments and innovations, and the question is whether it is going to coalesce into something bigger," explains **Andrew Hoffman,** a professor at the University of Michigan's School for Environment and Sustainability, who studies academic engagement in society. The environmental sciences lead the pack in this area, as do some biomedical and public health fields, he says.

Beyond PowerPoint

The University of Washington College of the Environment is aiming to serve as a model for scientist engagement and impact, says Dean **Lisa Graumlich.**

The college's midwife may be **Jane Lubchenco,** a prominent environmental scientist at Oregon State University in Corvallis. In the late 1990s, Lubchenco used her considerable influence to call on scientists to engage more broadly in society and help shape the public conversation on environmental issues. She later cofounded COMPASS and the Leopold Leadership cont.>

PHOTOS: (CLOCKWISE FROM TOP) © HXDBZXY/SHUTTERSTOCK.COM; KENNETH R. WEISS; BENJ DRUMMOND

POSTDOCTORAL FELLOWSHIPS IN GENOMIC BIOLOGY

AT THE UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

The Carl R. Woese Institute for Genomic Biology at the University of Illinois at Urbana-Champaign offers a number of fellowships for truly exceptional young scholars who have completed their PhD within the last several years, and are looking for a stimulating and supportive interdisciplinary environment to carry out independent and collaborative research in the field of genomic biology.

IGB Fellows will spend up to three years conducting research in one of several research themes in the institute, and ideally this research will also overlap with two or more of these thematic areas. The IGB also offers a Woese Fellowship for those at the forefront of their field in evolution and the emergence of life, or other rapidly developing areas of quantitative biology and genomics. Visit www.igb.illinois.edu/carl-r-woese-fellows for more information about the institute, the research themes and the application procedure. **The closing date for all positions is March 1, 2018. Fellows will be announced on or about April 1, 2018.**

ANTICANCER DISCOVERY FROM PETS TO PEOPLE

Seeking candidates with experience in anticancer drugs, drug discovery, veterinary medicine, computational comparative oncology and genomics in animal models.

BIOCOMPLEXITY

Seeking candidates with a strong quantitative background to join a variety of projects in: astrobiology and the emergence of life, geomicrobiology, metabolism and metagenomics of microbial communities, and the systems biology of microbes, biofilms and ecosystems.

BIOSYSTEMS DESIGN

Seeking candidates with experience in synthetic biology, metabolic engineering, protein engineering, genome engineering, systems biology, mass spectroscopy, laboratory automation, computer aided design, and bioinformatics.

COMPUTING GENOMES FOR REPRODUCTIVE HEALTH

Seeking candidates with experience in computational medicine, reproductive biology, statistical genomics, epigenomics, social science, and engineering and systems approaches to human health to develop predictive medicine.

GENE NETWORKS IN NEURAL & DEVELOPMENTAL PLASTICITY

Seeking candidates with strong bioinformatics skills and expertise in gene regulation (especially relating to transcription factor dynamics or epigenomics), evolutionary biology, neuroscience, neurogenomics, animal behavior, systems biology, and bioinformatics.

GENOMIC ECOLOGY OF GLOBAL CHANGE

Seeking candidates with a strong background in plant biology and a record of expertise in molecular biology, genomic ecology, synthetic biology, physiology, and modeling of gene networks or ecosystem function.

INFECTION GENOMICS FOR ONE HEALTH

Seeking candidates with experience in microbial ecosystems, antimicrobial resistance, infection biology, ecology, evolution, microbiology, virology, biomedical sciences, agricultural and food sciences, entomology, engineering, and anthropology.

MICROBIOME METABOLIC ENGINEERING

Seeking candidates with experience in microbial physiology, microbial biochemistry, microbial ecology, enzymology, nutrition, animal model development, toxicology and environmental health, and systems biology to understand the microbiome's role in health.

MINING MICROBIAL GENOMES

Seeking candidates with experience in microbially produced natural products, enzyme evolution, bacterial metabolism, bacterial genetics, molecular biology, biochemistry, enzyme evolution, metabolic engineering, organic synthesis, bioinformatics, and metagenomics.

OMICS NANOTECHNOLOGY FOR CANCER PRECISION MEDICINE

Seeking candidates with experience in nanotechnology, oncology, precision medicine, therapeutics, genomics, molecular and cellular biology, bioinformatics, engineering, bioengineering, and chemistry.

REGENERATIVE BIOLOGY & TISSUE ENGINEERING

Seeking candidates with experience in emerging topics in regenerative medicine and tissue engineering: biomaterials, (stem) cell biology, engineered cancer models, immunomodulation, microfluidics, tissue-on-a-chip, three-dimensional printing, and drug delivery.

ILLINOIS

 The Carl R. Woese Institute for Genomic Biology is an interdisciplinary institute dedicated to transformative research and technology in life sciences using team-based strategies to tackle grand societal challenges. Visit www.igb.illinois.edu for additional information.

"Something in between sitting ensconced in our ivory tower and playing more of an advocate and activist . . . is where we need to be."

– Andrew Hoffman

Program, a communications program for mid-career scientists, which Graumlich took part in during its first year.

When Graumlich became dean at the college's 2010 founding, she had absorbed Lubchenco's message and had developed the tools to help implement it. "This goes beyond beautiful PowerPoints," says Graumlich. "It informs everything we do."

The college brings together researchers as diverse as atmospheric scientists and fisheries students, who can all access communications resources such as intensive workshops. For example, a workshop specifically designed for graduate students typically fills its 20 slots within one day.

Like similar workshops elsewhere, the college's workshops focus on basic communications skills, teaching researchers how to define and understand their audience, develop a narrative, and come across as relatable and human.

Participants use the COMPASS message box, and also typically interact with a panel of journalists who challenge them to explain their research clearly, in front of their peers—an experience Kelly calls "intense."

Kelly explains that overcoming the need for approval from peers is a major communication barrier. When scientists broaden their audience, they often pare down their usual cautionary statements, eliminate jargon, and generate a simple message—all of which can aggravate an expert. "If you are communicating effectively you are only going to satisfy 95 percent of your peers," says Kelly. "You have to forget about that other 5 percent."

Researchers at UW can also take courses that teach specific skills such as making videos or graphics, or engage in a public-speaking program—and they can access one-on-one training to prepare for a media interview or a talk with a policymaker.

Kelly also recently moderated a panel discussion on communicating about socially and politically charged science topics. Graumlich adds that the emphasis on impact and engagement has also helped her college recruit young researchers drawn to this mission.

Other institutions offering programs to train scientists in communications include the University of Michigan, which has a workshop and community events, launched in 2013 by two graduate students. Cornell University and the University of Wisconsin have had such programs for decades, and dozens more are cropping up, some in the early stages of growth.

Homegrown and international programs

"At this point, most of the training that happens is sort of homegrown, and people are reinventing the wheel," says **John Besley,** who studies the field and also runs a communications program for scientists at the Department of Advertising and Public Relations, at Michigan State University in East Lansing. He notes that instructors do not all go to the same meetings or belong to the same science societies, so "no one knows what everyone else is doing."

Some homegrown programs are drawing inspiration from COMPASS or workshops at the American Association for the Advancement of Science (AAAS, the home of *Science* Careers). Another major player is the Alan Alda Center for Communicating Science, at Stony Brook University in New York. The center focuses on improvisation and listening skills to help scientists and engineers be more relatable—a goal of its founder, Alan Alda, former host of the PBS show *Scientific American Frontiers*. The center has trained nearly 10,000 people across the country, according to director **Laura Lindenfeld.**

Internationally, the London-based Science Media Centre hosts workshops throughout the United Kingdom, and has inspired similar endeavors in Germany, Canada, and other countries. In Australia, most universities have robust communications programs, says **Joan Leach,** who leads the Australian National Centre for the Public Awareness of Science (CPAS), at the Australian National University in Canberra. CPAS has partnerships with the Alan Alda Center as well as institutions in Indonesia, Africa, and New Zealand.

But what should a scientist do if their institution provides little or no communications support? **Brian Lin,** who oversees the AAAS-hosted media portal EurekAlert!, runs communications workshops in Japan and China, countries that are just beginning to build public relations offices at their universities, akin to what happened in the United States about 15 years ago. He advises researchers with a hot paper on deck to contact the journal publishing their study for guidance. People can also hire the Alan Alda Center, COMPASS, or other individual trainers for one-on-one sessions before facing a media storm.

Some scientific societies, such as Newswise from the American Society for Cell Biology, also offer training at meetings as well as other support. **Erin Wirth,** for instance, fielded media inquiries about a study on earthquake hazards while she was a postdoc at the University of Washington. A press officer there directed Wirth, now at the U.S. Geological Survey in Seattle, to the website of the American Geophysical Union, where she found a worksheet similar to the COMPASS message box. Wirth's initial media exposure has led her to an outreach opportunity speaking at a public forum in a nearby town.

The science of science communications

Communications training programs are increasingly leveraging research about best practices for scientists, says Besley, such as the value of being relatable and telling a story. And they are helping researchers use their time wisely.

To keep researchers focused, more programs are encouraging scientists to identify their communications goals and audience. After that, scientists can choose an activity they feel comfortable with, such as writing for an outlet like cont.>

PHOTO: BLOOMBERG NEW ENERGY FINANCE, FUTURE OF ENERGY SUMMIT

83rd Cold Spring Harbor Laboratory Symposium on Quantitative Biology

May 30 - June 4, 2018

Poster Abstract Deadline: March 31, 2018

Brains & Behavior

Order and Disorder in the Nervous System

Arranged by: David Stewart and Bruce Stillman - Cold Spring Harbor Laboratory

Cellular Basis of Neurodegeneration ✦ Hormonal Regular and Mood Disorders
Cortico-Striatal Circuits in Perception and Action ✦ Memory Systems and Aging
Computational Psychiatry ✦ New Techniques for Human Brain Manipulation
Deep Brain Stimulation for Psychiatry ✦ Neural Circuit of Anxiety
Sleep and Sleep Disorders ✦ Development of Social Cognition ✦ Restoring Vision
Addicted Brain: Circuits and Treatment Frontiers ✦ Brain Machine Interface

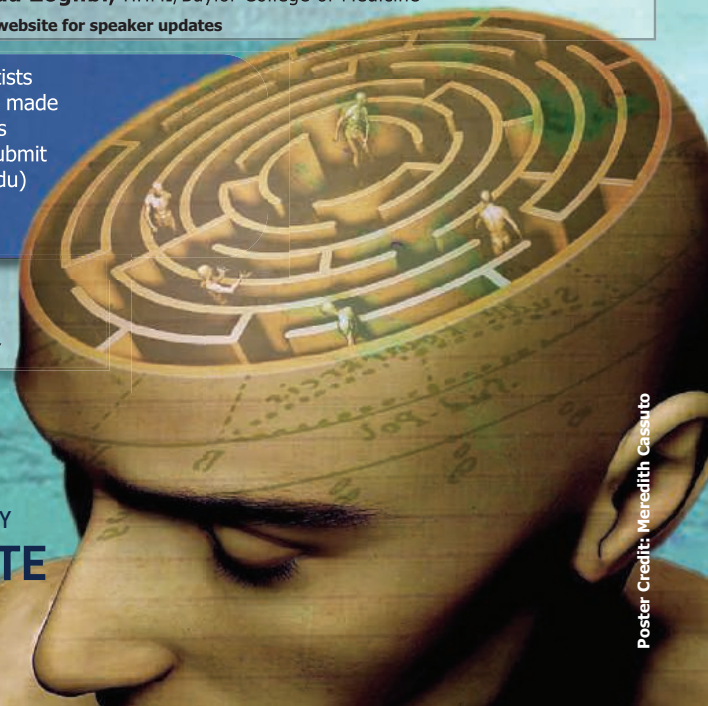
Invited Speakers

David Anderson, HHMI/Caltech
Hugo Bellen, HHMI/Baylor College of Medicine
Antonello Bonci, National Institute on Drug Abuse
Edward Boyden, Massachusetts Institute of Technology
Kim Butts Pauly, Stanford University
Pico Caroni, Friedrich Miescher Institute SWITZERLAND
Anne Churchland, Cold Spring Harbor Laboratory
Yang Dan, University of California, Berkeley
Karl Deisseroth, Stanford University
Ricardo Dolmetsch, Novartis
Emad Nader Eskandar, Harvard Medical School
Frances Edwards, University College London, UK
Guoping Feng, Massachusetts Institute of Technology
Paul Glimcher, New York University
Michel Goedert, MRC Laboratory of Molecular Biology, UK
Joshua Gordon, National Institute of Mental Health
Seth Grant, University of Edinburgh, UK
Hailan Hu, Zhejiang University School of Medicine, China
Richard Huganir, Johns Hopkins University School of Medicine
Michael Kahana, University of Pennsylvania
Eric Kandel, HHMI/Columbia University
Adam Kepecs, Cold Spring Harbor Laboratory
Bo Li, Cold Spring Harbor Laboratory
Stephen Liberles, Harvard Medical School
Andres Lozano, University of Toronto, Canada
Christian Luscher, University of Geneva, Switzerland
Zachary Mainen, Champalimaud Neuroscience Program, Portugal
Robert Malenka, Stanford University School of Medicine
Steven McKnight, UT Southwestern Medical Center
Istvan Mody, UCLA School of Medicine
P. Read Montague, Virginia Tech Carilion Research Institute
Eric Nestler, Mount Sinai University
Michael Platt, University of Pennsylvania
Clifford Saper, Harvard Medical School
Akira Sawa, Johns Hopkins University
Daniela Schiller, Icahn School of Medicine at Mount Sinai
Amita Sehgal, The University of Pennsylvania
Nirao Shah, Stanford University
Scott Small, Columbia University
Beth Stevens, Harvard Medical School; Children's Hospital, Boston
Kevin Tracey, Feinstein Institute for Medical Research
Li-Huei Tsai, Massachusetts Institute of Technology
Kay Tye, Massachusetts Institute of Technology
Naoshige Uchida, Harvard University
Flora Vaccarino, Yale University
Linda Wilbrecht, University of California, Berkeley
Daniel Wolpert, University of Cambridge, UK
Mehmet Fatih Yanik, ETH Zürich Switzerland
Tony Zador, Cold Spring Harbor Laboratory
Huda Zoghbi, HHMI/Baylor College of Medicine

See website for speaker updates

Ten **Tianqiao & Chrissy Chen Fellowships** will be awarded to young scientists presenting their latest research at the 2018 Symposium. Award decisions will be made on the basis of scientific merit and approach by CSHL and TCCI. Young scientists (graduate students, postdocs and junior faculty) interested in applying should submit abstracts and a brief one-paragraph resume/bio to Val Pakaluk (pakaluk@cshl.edu) by March 31, 2018. Chen Fellows receive a stipend covering the full academic or student attendance package.

Join us for the East Coast premiere of a documentary created by the Tianqiao & Chrissy Chen Institute to highlight recent developments and outline the exciting future of brain science and health.
Screenings: May 29 & June 2; @7.00pm; Grace Auditorium, Cold Spring Harbor Laboratory



Poster Credit: Meredith Cassuto

Cold Spring Harbor Laboratory is proud to partner with The Tianqiao & Chrissy Chen Institute to bring you the 2018 Symposium

Cold Spring Harbor Laboratory
MEETINGS & COURSES PROGRAM

CHEN TIANQIAO & CHRISSY INSTITUTE

Registration, poster abstract submission & further information:
meetings.cshl.edu/symp18 meetings@cshl.edu (516) 367-8346

Featured participants

Alan Alda Center for Communicating Science
www.aldacenter.org

American Geophysical Union
news.agu.org

Australian National Centre for the Public Awareness of Science, Australian National University
cpas.anu.edu.au

College of the Environment, University of Washington
environment.uw.edu

COMPASS
www.compassscicomm.org

Department of Advertising and Public Relations, Michigan State University
comartsci.msu.edu/advertising-public-relations

EurekAlert!
www.eurekalert.org

Jane Lubchenco, Oregon State University
gordon.science.oregonstate.edu/lubchenco/jlcvbio

Lund University
www.lunduniversity.lu.se

Marcus Autism Center
marcus.org/about-us

Newswise
www.newswise.com/institutions/newsroom/979

School for Environment and Sustainability, University of Michigan
seas.umich.edu

School of Marine and Environmental Affairs, University of Washington
smea.uw.edu

Science Media Centre
www.sciencemediacentre.org

The Conversation (theconversation.com), which features experts in various fields, or doing their own podcast. Scientists are also increasingly using Twitter, particularly to reach peers about research that interests them. Says Besley, "People are becoming increasingly sophisticated about what they need to do to achieve their goals."

Though everyone can improve, Hoffman says that broader communication and engagement is not everyone's strength. And researchers need to think carefully about their role in any conversation—a sentiment echoed by Kelly, who notes that the scientific community does not view excess self-promotion favorably.

"There is something in between sitting ensconced in our ivory tower and totally jumping in and playing more of an advocate and activist–which is dangerous," says Hoffman. "Somewhere in between is where we need to be, but I think that place is always changing."

Building a network

The direct influence of COMPASS and similar programs is hard to trace, but their reach is international in scope. When **Maren Wellenreuther** was a postdoc at Lund University in Sweden in 2013, she relied on Baron's book to help build a communications program for scientists. Wellenreuther knew very little starting out, but *Escape from the Ivory Tower* still serves as a guide for the week-long course, which draws Ph.D. students throughout Scandinavia and Northern Europe.

Wellenreuther now has a position in New Zealand, at Plant and Food Research in Nelson, where she is planning her next move in the communications realm. "The younger generation especially realizes that, while [training in science communications] takes time away from their course of study, it also gives back so much to you and makes you a better scientist."

At Lund University, the communications course is cotaught by **Daniel Conley,** an aquatic scientist who is also indirectly part of Nancy Baron's network. Inspired by the Leopold Fellowship, Conley began a similar program in 2011 focused on marine sciences, the Vega Fellows program. And the network that began there spread outwards—two Vega Fellows consulted during their train trip back from the program and started Baltic Eye, an organization that works with journalists on issues affecting the Baltic Sea.

Conley echoes the motivations of many scientists who undertake similar endeavors. "I want to make a difference," he says.

Providing incentives

Both Conley and Wellenreuther feel that Lund University was supportive of their work in science communications. One factor contributing to this, says Conley, is that some Swedish funding agencies require researchers to engage with the wider community. Such requirements are helping to bolster similar endeavors worldwide.

In the United Kingdom, national requirements for impact and engagement encourage communications efforts at research institutions, says **Fiona Fox,** head of the Science Media Centre, founded in 2002 in part as a response to public misunderstanding about the safety of genetically modified organisms.

In the United States, grant proposals submitted to the U.S. National Science Foundation are evaluated for their broader impact on society, and some grants from the U.S. National Institutes of Health (NIH) also have an outreach component. **Chris Gunter,** for instance, is a principal investigator for the outreach component of a large NIH Autism Center of Excellence grant that funds the Marcus Autism Center in Atlanta, Georgia. As director of communications operations there, she is planning events at the Atlanta Science Festival, including a discussion of the portrayal of autism in the media. In addition, she coleads a three-hour media training workshop for the faculty at the center.

Offering rewards

While many scientists may wish to engage a broader audience, they may feel hamstrung by the reward system at their institution. Few institutions have formal systems to recognize or reward such endeavors through tenure or promotion decisions.

Fortunately, there are new efforts underway to change that. The Mayo Clinic, for instance, now includes social media involvement in its criteria for academic advancement. And the University of Michigan Ross School of Business, where Hoffman studies, now includes a category called "practice'' in its annual review, which can include media engagement. At UW College of the Environment, the impact of engagement and communications activities is evaluated under "service," a promotion category present at many universities.

Ryan Kelly will be assembling his tenure and promotion package this year. He will highlight his outreach activities with the media and policymakers, along with his role as an advisor to the Public Comment Project, a web portal that facilitates public comment on proposed federal regulations. He feels that these activities will be viewed positively.

And at UW, the dean is on board. "Part of being a member of the faculty at the college is to be an excellent scholar and then to strategically share that scholarship in a way that has impact," says Graumlich. "This is not some kind of icing on the cake. We are changing the culture [of science]."

Charlotte Schubert is a freelance journalist based in Seattle.

Stony Brook University

Multiple Postdoctoral Positions Spring/Summer 2018

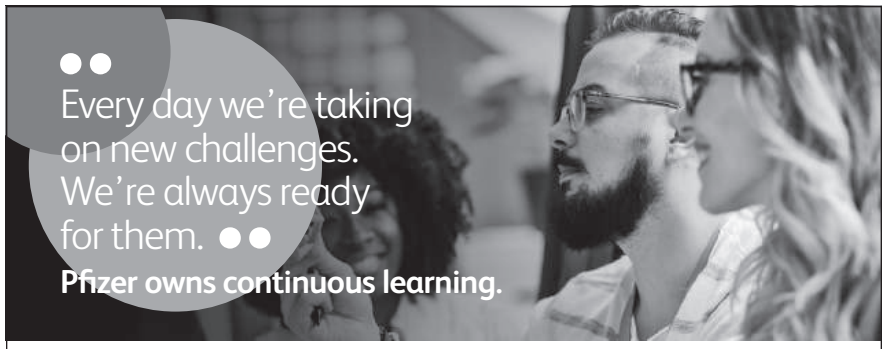
Stony Brook University is recruiting for multiple postdoctoral positions in various subspecialties, for the upcoming spring and summer months.

Stony Brook has been characterized by innovation, energy and progress, and making groundbreaking discoveries since its beginning half a century ago.

Any interested candidates are invited to visit our JOBS page.

www.stonybrook.edu/postdocjobs

Stony Brook University/SUNY is an equal opportunity, affirmative action employer.



Pfizer Worldwide Research and Development Postdoctoral Program

At Pfizer, postdocs are trained to become successful, independent investigators, capable of formulating and addressing important scientific hypotheses. In addition, trainees receive broad exposure to the process of drug discovery, from idea to clinical trials. Areas of scientific focus include cardiovascular and metabolic diseases, comparative medicine, drug safety, biotherapeutics/protein engineering, inflammation and immunology, medicinal chemistry, oncology, pharmacology, vaccines, and clinical, computational, and genomic sciences.

We recruit highly motivated Ph.D. recipients with an outstanding track record of scientific productivity and a passion for ground-breaking, fast-paced research that facilitates the development of innovative therapies for human diseases. Our program promotes dissemination of research results through publications and participation in scientific meetings, provides opportunities for collaboration with leading academic labs and industry consortia, and offers exceptional professional development training and networking opportunities.

To explore our program and research, visit us online at:
www.pfizer.com/careers/en/postdoctoral-program

pfizercareers.com

Pfizer OWNIT!

Cleveland Clinic
Lerner Research Institute

Postdoctoral Fellowships Available

The Lerner Research Institute is home to all basic, translational and clinical research at Cleveland Clinic, the No. 2 ranked U.S. hospital by *U.S. News and World Report*. With over $140M in federal grants and an annual research budget of $260M, the Lerner is consistently ranked among the top research institutes in the nation. Postdoctoral fellows routinely obtain grant funding and first-author publications in top-tier journals.

Postdoctoral Job Opportunities:
http://www.lerner.ccf.org/jobs/postdoctoral/
For further information email: RETC@ccf.org

The Lerner Experience

- Opportunity to train among world-class scientists and physician-scientists in a top-ranked healthcare institution
- Multidisciplinary, disease-focused research programs
- 175 principal investigators in 12 departments with over 700,000 square feet of research space
- Competitive salary and benefits package
- Active Postdoctoral Association and 250+ postdocs
- Career development opportunities and support through the Research Education and Training Center

City of Cleveland

Cleveland is a multicultural city with nationally acclaimed museums, sports, restaurants, and music and arts programs. Situated on Lake Erie, the area offers stunning views, beaches, and water sports. Low cost of living, with below average traffic and commute times for major cities.

Neuroscience Post-doctoral Programme

Linköping University is one of Sweden's six large universities, currently enrolling 27,000 students. The Centre for Systems Neurobiology involves some 50 independent research groups, from the Faculty of Medicine and the Faculty of Science and Engineering, as well as the University Hospital.

The Centre for Systems Neurobiology is now seeking Postdoctoral Fellows within several neuroscience research areas: Addiction, Animal Behavior, Electrophysiology and Circuits, Neuroimaging, Neurodegeneration, Neuroendocrinology, Neurodevelopment, Pain, Psychiatry, and Sensory Systems. Applications will be reviewed continuously. For full consideration please apply as soon as possible.

LINKÖPING UNIVERSITY

For more details regarding the Centre, the different research labs involved in the programme, and to submit a letter-of-intent please go to: liu.se/en/research/center-for-systems-neurobiology. For information regarding the university and the region, please go to: liu.se, eastsweden.com.

BIODISCOVERY INSTITUTE UNT

BIODISCOVERY INSTITUTE, UNIVERSITY OF NORTH TEXAS TWO ASSISTANT/ASSOCIATE PROFESSOR POSITIONS

The University of North Texas (UNT), Denton, Texas, invites applications from outstanding scientists for two academic-year positions at the assistant or associate professor levels in the BioDiscovery Institute (BDI), one of four recently established Institutes of Research Excellence. Candidates addressing important fundamental and applied questions in the synthesis and development of bio-based products for agriculture, materials engineering, bioenergy or health benefits are especially encouraged to apply. We are seeking individuals working in two general areas: (1) microbial metabolism and synthetic biology and (2) structure-based enzyme design. These new hires will complement and interact with existing researchers in systems modeling of metabolism, plant biochemistry, and biotechnology. Dedicated, newly renovated research space, competitive start-up packages, and state-of-the-art core facilities in genomics and metabolomics will support these new members and their colleagues in BDI.

Successful candidates will bring an internationally recognized research program, interact with diverse faculty, staff and students within BDI and across UNT, and contribute to graduate and/or undergraduate training. Applicants should have a Ph.D. in a biological, computational, physical or engineering sciences discipline, an established record of funding and publications in one of the target areas described above, and a commitment to excellence in research within an interdisciplinary environment. Applications should be made through the UNT Academic Resources website (**https://facultyjobs.unt.edu**) and should include a cover letter, curriculum vitae, two-page summary of research interests and accomplishments, one-page statement of teaching philosophy, and the names and contact information for three references. Questions may be directed to Brian G. Ayre (**bgayre@unt. edu**), Search Committee Chair, or Kent D. Chapman (**chapman@unt.edu**), BDI Director. Screening of applications will begin immediately and will continue until the positions are closed. For best consideration, apply by **February 15, 2018**.

The University of North Texas is an Equal Opportunity/Access/Affirmative Action/Pro Disabled and Veteran Institution committed to diversity in its employment and educational programs, thereby creating a welcoming environment for everyone.

AUGUSTA UNIVERSITY MEDICAL COLLEGE OF GEORGIA

ASSISTANT/ASSOCIATE PROFESSOR/PROFESSOR NEUROSCIENCE/NEUROANATOMY EDUCATOR

The Department of Neuroscience and Regenerative Medicine in collaboration with the office of Academic Affairs of the Medical College of Georgia at Augusta University, Augusta, GA, invites applications for a full-time faculty position at the rank of Assistant/Associate/Full Professor. The successful candidate will play a principal educator role and work with a team of colleagues in teaching clinically oriented neuroscience and neuroanatomy within an integrated pre-clinical curriculum in the Medical College of Georgia and neuroscience in the College of Graduate Studies. He/she will have a partial assignment to the Education Innovation Institute and have opportunity to participate in educational research, faculty development, or both. The Department of Neuroscience and Regenerative Medicine presently includes 15-tenured/tenure track core faculty with significant research accomplishments and educational excellence in neural development, neural cell and molecular biology, and disorders of the peripheral and central nervous system. Candidates must have a PhD and/or MD and have teaching experience, two years preferred, in medical neuroscience. Salary is dependent on qualifications and experience. Augusta University is a state-supported comprehensive academic medical center whose mission is to train physicians and other health professionals to meet the health care needs of the state.

Applicants should submit a letter with description of teaching interests and experience, curriculum vitae, and names of three references to Dr. Richard Cameron, c/o Deenie Cerasuolo (dcerasuo@augusta.edu). Please also apply at **http://www.augusta.edu/hr/jobs/faculty/** position **#12720**. Review of applications will begin immediately and continue until the position is filled.

Augusta University is an Equal Employment Opportunity and Equal Access Institution.

POSTDOCTORAL OPPORTUNITIES

MOTE MARINE LABORATORY & AQUARIUM

2019-2020 POSTDOCTORAL RESEARCH FELLOWSHIPS

Mote Marine Laboratory announces two new two-year positions beginning sometime between October 2018 and January 2019, to support independent investigators based in Sarasota, Florida USA. Applications are invited from recent (since December 2015) Ph.D. graduates, including those with firm expectation of graduation by December 2018. However, at time of appointment, a doctoral degree must have been awarded. Proposals for any field of marine research are invited.

However the following fields are especially encouraged:

- **Marine Ecotoxicology:** Source, fate and effects of toxic chemicals in the coastal marine environment, including analysis of harmful algal toxins, pesticides and other endocrine disrupting compounds, focusing on synergistic effects from exposure to multiple stressors, and identifying biomarkers of sublethal effects manifested in subsequent generations.
- **Sirenian biology and conservation:** Anatomy, morphology and physiology, reproduction and mating strategies, sensory biology, cognition and behavior, behavioral ecology, habitat use, spatial ecology, and genetics of Florida manatees.

Competitive applications will focus on research programs that are relevant to conservation and sustainable uses of marine biodiversity, healthy habitats, and natural resources; will bring or propose new multi-investigator/institutional collaborations to Mote, and will be cognizant of global issues. A salary of $ 49,000 and funding for research start-up and expenses are provided. See **https://mote.org/about-us/employment-opportunities** for details and application instructions. The deadline for applications is **August 31, 2018** and finalists will be announced in October.

All qualified applicants will receive consideration for employment without regard to race, color, religion, sex, sexual orientation, gender identity, national origin, disability, protected veteran status or other protected category. Mote is an E-Verify Employer.

By **Kevin F. Boehnke**

Cheating on my mentor

For the first 2 years of my Ph.D. program, my primary adviser was always available when I needed help, promptly responding to emails and meeting with me when questions arose. But that abruptly changed when he went on sabbatical and left the country. My emails were rarely answered, and our scheduled meetings were often canceled. I did what I could to push my research forward on my own, but I felt I was floundering. I grew frustrated and uncertain about my future, even questioning whether I wanted to pursue a research career. After sufficient wallowing in self-pity, however, I decided to take matters into my own hands by seeking the support that I needed.

Haphazardly, I started building a broader network of mentors. I reached out to collaborators, committee members, teachers, and friends to ask for advice about my research and career development. It was nerve-wracking to call on people I didn't know well or hadn't spoken to about these topics before. I didn't always get responses, nor were the responses I received always helpful. But several patient people listened, gave judicious advice, and connected me to others with whom I had shared interests. Others generously gave their time to help me develop side projects and novel hypotheses, offering suggestions based on their expertise. These relationships and experiences lifted my spirits, broadened my perspective, and reignited my interest in research.

"Haphazardly, I started building a broader network of mentors."

After being away for several months, my primary adviser re-emerged. I was excited to have his insights again, but I feared that he might not approve of the additional mentoring I had sought out. However, instead of being upset that I had "cheated" on him, he was pleased. He recognized that the new perspectives added value and context to our work, and he encouraged me to continue.

From that point on, I deliberately pursued this team mentorship model. My advisers, people with varied perspectives and expertise, didn't necessarily work together or even meet. Nonetheless, they were my team in the sense that they supported my growth as a public health researcher. For example, one adviser encouraged me to participate in a workshop about computational methods for policy-related research, which at the time was outside my comfort zone. I came away with a better understanding of how to design research for maximum real-world impact, which I have since used to shape my studies. Overall, the support I received from my team of mentors helped me become more confident in pursuing my own ideas, which ultimately helped me find deep satisfaction in my Ph.D. work. When I moved on to my postdoc, I sought out and found a research group that explicitly prioritizes team science.

Building and maintaining relationships with several mentors rather than one takes time and adds complexity to my life. During my Ph.D. work, for example, negotiating reasonable milestones with multiple mentors required more effort up front, and some of my proposed experiments took far longer to design as I assimilated advice from experts in a variety of disciplines. However, these front-loaded costs paid off. My research was both more impactful scientifically and more meaningful to me personally. Moreover, I learned to communicate across disciplines and get people with different viewpoints on the same page—skills that are at the core of my work in public health.

Many students I know refer to their mentor as an individual with a larger-than-life persona, the arbiter of success and failure. This isn't necessarily a bad thing, but it can place undue strain on both the mentor and the student, leading to frustration and power struggles. As I learned firsthand, it is not realistic to always expect immediate feedback, exhaustive answers, and quick and productive meetings all from a single individual, who is, after all, human, with their own life and career to foster. With my mentoring team, I receive different things from different people, and incorporating their perspectives into my work has enriched both my research and my life. ■

Kevin F. Boehnke is a postdoctoral research fellow at the University of Michigan in Ann Arbor. *He is indebted to his many mentors. Do you have an interesting career story? Send it to SciCareerEditor@aaas.org.*

ILLUSTRATION: ROBERT NEUBECKER